Multiple Random Variables

Joint Probability Density

Let X and Y be two random variables. Their **joint distribution function** is $F_{XY}(x,y) \equiv P[X \le x \cap Y \le y]$.

$$0 \le F_{XY}(x,y) \le 1 \quad , \quad -\infty < x < \infty \quad , \quad -\infty < y < \infty$$

$$F_{XY}(-\infty,-\infty) = F_{XY}(x,-\infty) = F_{XY}(-\infty,y) = 0$$

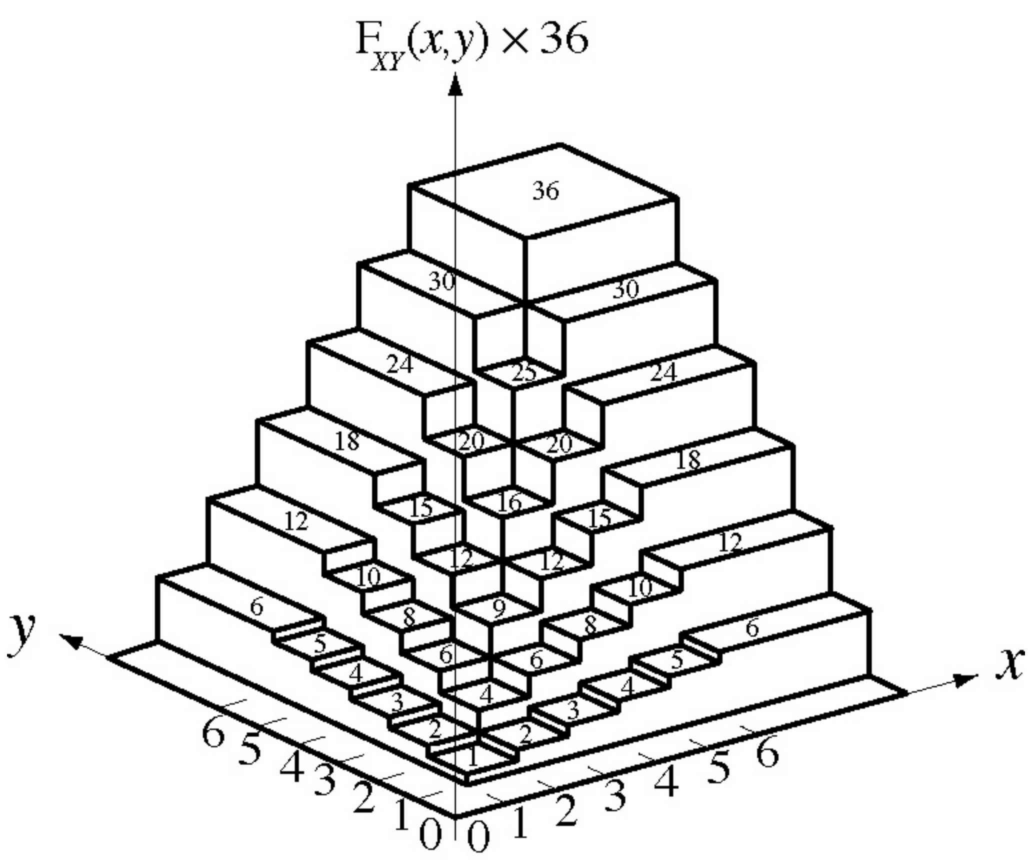
$$F_{XY}(\infty,\infty) = 1$$

$F_{XY}(x,y)$ does not decrease if either x or y increases or both increase

$$F_{XY}(\infty,y) = F_Y(y) \text{ and } F_{XY}(x,\infty) = F_X(x)$$

Joint Probability Density

Joint distribution function for tossing two dice



Joint Probability Density

$$\mathrm{f}_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y}\left(\mathrm{F}_{XY}(x,y)\right)$$

$$\mathrm{f}_{XY}(x,y) \geq 0 \quad , \quad -\infty < x < \infty \quad , \quad -\infty < y < \infty$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{f}_{XY}(x,y)\,dx\,dy = 1 \qquad \mathrm{F}_{XY}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} \mathrm{f}_{XY}(\alpha,\beta)\,d\alpha\,d\beta$$

$$\mathrm{f}_X(x) = \int_{-\infty}^{\infty} \mathrm{f}_{XY}(x,y)\,dy \quad \text{and} \quad \mathrm{f}_Y(y) = \int_{-\infty}^{\infty} \mathrm{f}_{XY}(x,y)\,dx$$

$$\mathrm{P}\left[x_1 < X \leq x_2 \ , \ y_1 < Y \leq y_2\right] = \int_{y_1}^{y_2}\int_{x_1}^{x_2} \mathrm{f}_{XY}(x,y)\,dx\,dy$$

$$\mathrm{E}\left(\mathrm{g}(X,Y)\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{g}(x,y)\,\mathrm{f}_{XY}(x,y)\,dx\,dy$$

Combinations of Two Random Variables

Example

X and Y are independent, identically distributed (i.i.d.) random variables with common PDF

$$\mathrm{f}_X(x) = e^{-x}\,\mathrm{u}(x) \quad \mathrm{f}_Y(y) = e^{-y}\,\mathrm{u}(y)$$

Find the PDF of $Z = X / Y$.

Since X and Y are never negative, Z is never negative.

$\mathrm{F}_Z(z) = \mathrm{P}\left[X / Y \le z\right] \Rightarrow \mathrm{F}_Z(z) = \mathrm{P}\left[X \le zY \cap Y > 0\right] + \mathrm{P}\left[X \ge zY \cap Y < 0\right]$

Since Y is never negative $\mathrm{F}_Z(z) = \mathrm{P}\left[X \le zY \cap Y > 0\right]$

Combinations of Two Random Variables

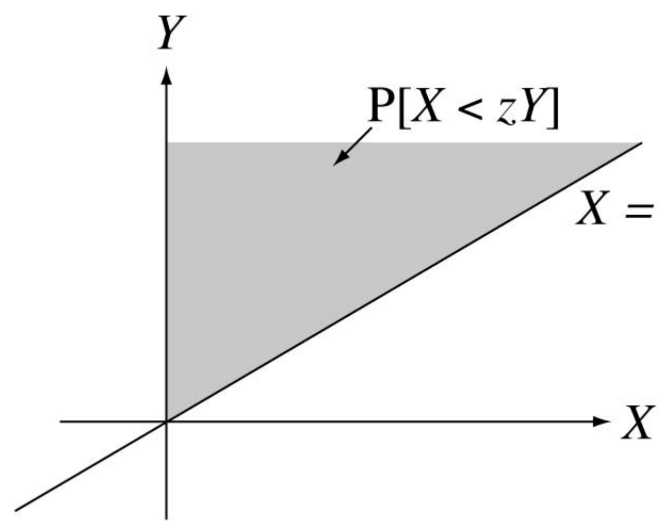
$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{zy} f_{XY}(x,y)\,dxdy = \int_0^{\infty} \int_0^{zy} e^{-x} e^{-y}\,dxdy$$

Using Leibnitz's formula for differentiating an integral,

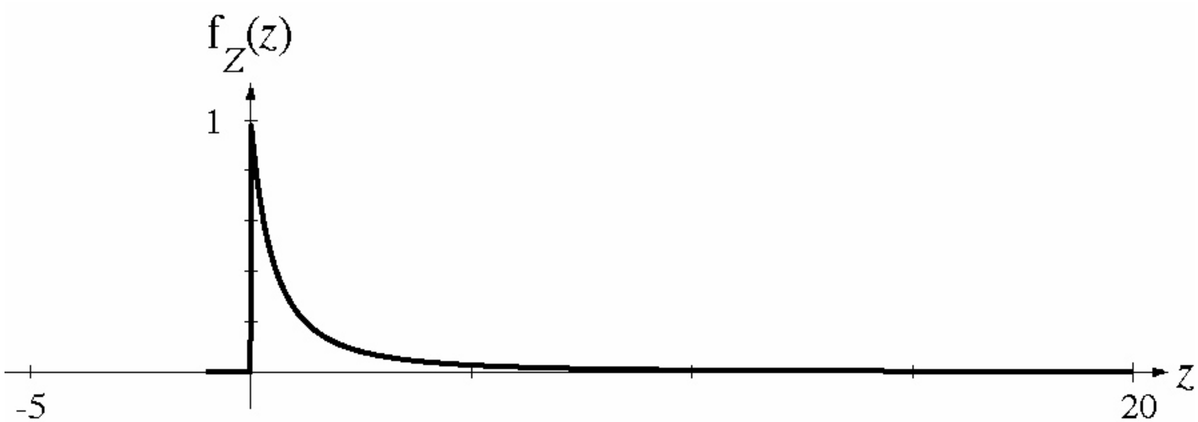
$$\frac{d}{dz}\left[\int_{a(z)}^{b(z)} g(x,z)\,dx\right] = \frac{db(z)}{dz} g(b(z),z) - \frac{da(z)}{dz} g(a(z),z) + \int_{a(z)}^{b(z)} \frac{\partial g(x,z)}{\partial z}\,dx$$

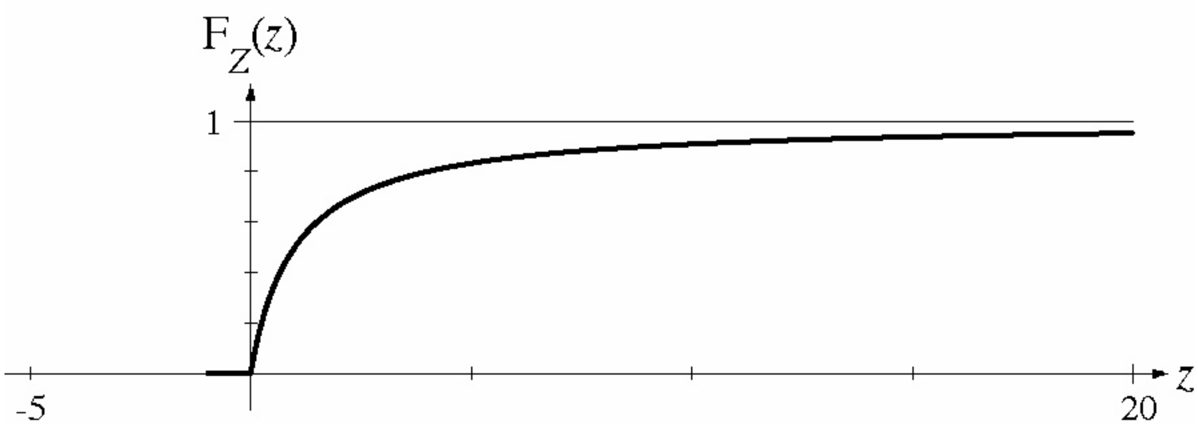
$$f_Z(z) = \frac{\partial}{\partial z} F_Z(z) = \int_0^{\infty} y e^{-zy} e^{-y}\,dy \quad , \quad z > 0$$

$$f_Z(z) = \frac{u(z)}{(z+1)^2}$$



Combinations of Two Random Variables





Combinations of Two Random Variables

Example

The joint PDF of X and Y is defined as

$$f_{XY}(x,y) = \begin{cases} 6x & , \ x \geq 0, y \geq 0, x + y \leq 1 \\ 0 & , \ \text{otherwise} \end{cases}$$

Define $Z = X - Y$. Find the PDF of Z.

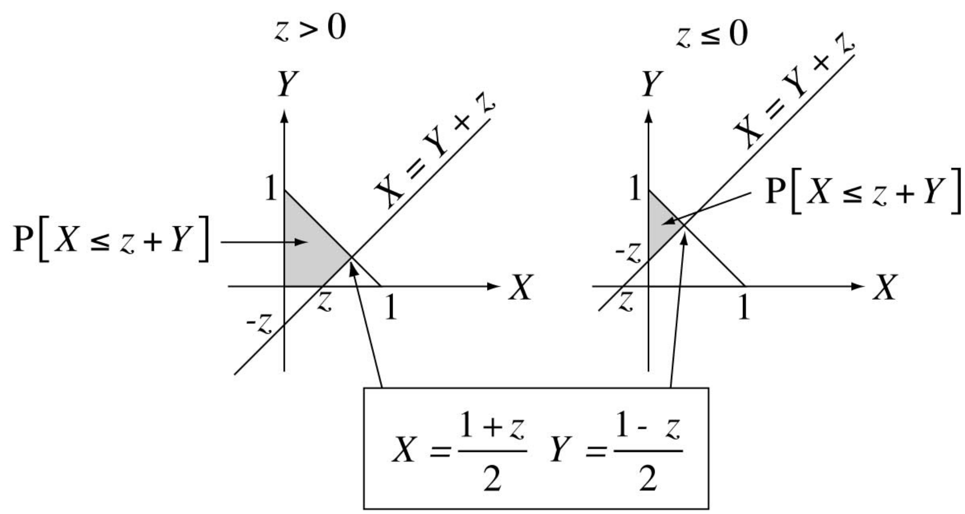
Combinations of Two Random Variables

Given the constraints on X and Y, $-1 \le Z \le 1$.

$Z = X - Y$ intersects $X + Y = 1$ at $X = \dfrac{1+Z}{2}$, $Y = \dfrac{1-Z}{2}$

For $0 \le z \le 1$, $\mathrm{F}_Z(z) = 1 - \displaystyle\int_0^{(1-z)/2} \int_{y+z}^{1-y} 6x\,dx\,dy = 1 - \int_0^{(1-z)/2} \left[3x^2\right]_{y+z}^{1-y} dy$

$\mathrm{F}_Z(z) = 1 - \dfrac{3}{4}(1-z)(1-z^2) \Rightarrow \mathrm{f}_Z(z) = \dfrac{3}{4}(1-z)(1+3z)$

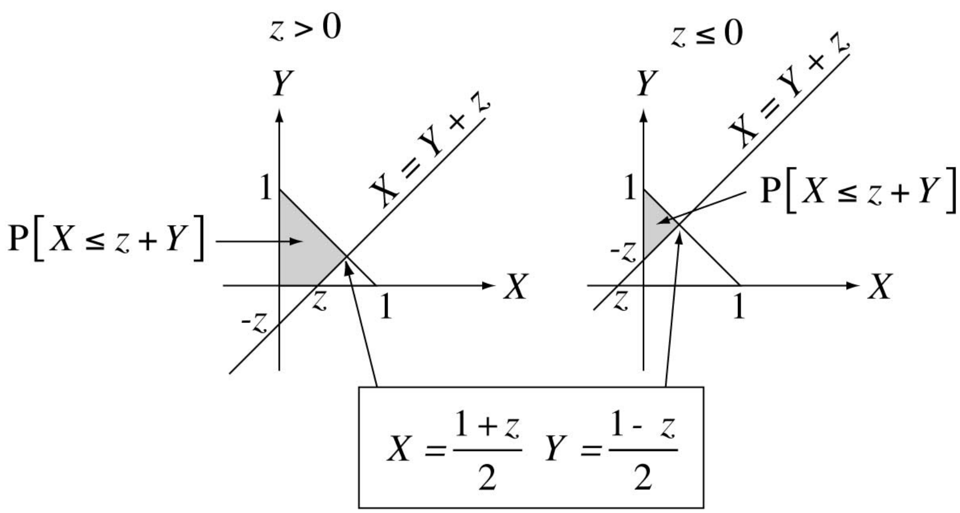


Combinations of Two Random Variables

For $-1 \le z \le 0$

$$\mathrm{F}_Z(z) = 2\int_{-z}^{(1-z)/2}\int_0^{y+z} 6x\,dx\,dy = 6\int_{-z}^{(1-z)/2}\left[x^2\right]_0^{y+z} dy = 6\int_{-z}^{(1-z)/2}(y+z)^2\,dy$$

$$\mathrm{F}_Z(z) = \frac{(1+z)^3}{4} \Rightarrow \mathrm{f}_Z(z) = \frac{3(1+z)^2}{4}$$



Joint Probability Density

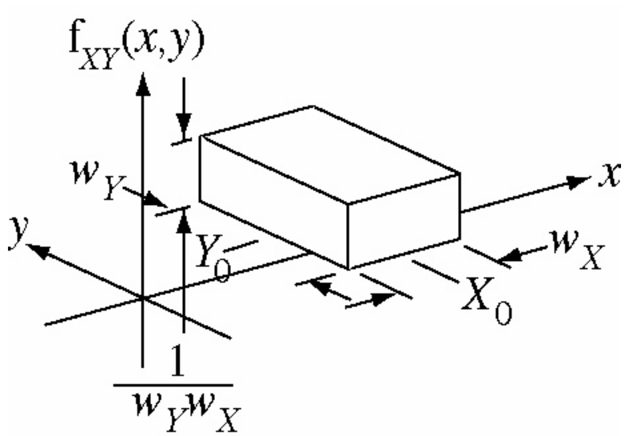
$$\text{Let } \mathrm{f}_{XY}(x,y) = \frac{1}{w_X w_Y}\,\mathrm{rect}\left(\frac{x-X_0}{w_X}\right)\mathrm{rect}\left(\frac{y-Y_0}{w_Y}\right)$$

$$\mathrm{E}(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,\mathrm{f}_{XY}(x,y)\,dxdy = X_0$$

$$\mathrm{E}(Y) = Y_0$$

$$\mathrm{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,\mathrm{f}_{XY}(x,y)\,dxdy = X_0 Y_0$$

$$\mathrm{f}_X(x) = \int_{-\infty}^{\infty} \mathrm{f}_{XY}(x,y)\,dy = \frac{1}{w_X}\,\mathrm{rect}\left(\frac{x-X_0}{w_X}\right)$$



Joint Probability Density

Conditional Probability $\quad \mathrm{F}_{X|A}(x)=\dfrac{\mathrm{P}\left[\left(X\le x\right)\cap A\right]}{\mathrm{P}\left[A\right]}$

Let $A=\left\{Y\le y\right\}$

$$\mathrm{F}_{X\,|\,Y\le y}(x)=\frac{\mathrm{P}\left[X\le x\cap Y\le y\right]}{\mathrm{P}\left[Y\le y\right]}=\frac{\mathrm{F}_{XY}(x,y)}{\mathrm{F}_{Y}(y)}$$

Let $A=\left\{y_1<Y\le y_2\right\}$

$$\mathrm{F}_{X\,|\,y_1<Y\le y_2}(x)=\frac{\mathrm{F}_{XY}(x,y_2)-\mathrm{F}_{XY}(x,y_1)}{\mathrm{F}_{Y}(y_2)-\mathrm{F}_{Y}(y_1)}$$

Joint Probability Density

Let $A = \{Y = y\}$

$$\mathrm{F}_{X\,|\,Y=y}(x) = \lim_{\Delta y \to 0} \frac{\mathrm{F}_{XY}(x, y+\Delta y) - \mathrm{F}_{XY}(x,y)}{\mathrm{F}_Y(y+\Delta y) - \mathrm{F}_Y(y)} = \frac{\frac{\partial}{\partial y}\left(\mathrm{F}_{XY}(x,y)\right)}{\frac{d}{dy}\left(\mathrm{F}_Y(y)\right)}$$

$$\mathrm{F}_{X\,|\,Y=y}(x) = \frac{\frac{\partial}{\partial y}\left(\mathrm{F}_{XY}(x,y)\right)}{\mathrm{f}_Y(y)} \;,\; \mathrm{f}_{X|Y=y}(x) = \frac{\partial}{\partial x}\left(\mathrm{F}_{X\,|\,Y=y}(x)\right) = \frac{\mathrm{f}_{XY}(x,y)}{\mathrm{f}_Y(y)}$$

Similarly $\mathrm{f}_{Y|X=x}(y) = \dfrac{\mathrm{f}_{XY}(x,y)}{\mathrm{f}_X(x)}$

Joint Probability Density

In a simplified notation $f_{X|Y}(x) = \dfrac{f_{XY}(x,y)}{f_Y(y)}$ and $f_{Y|X}(y) = \dfrac{f_{XY}(x,y)}{f_X(x)}$

Bayes' Theorem $f_{X|Y}(x) f_Y(y) = f_{Y|X}(y) f_X(x)$

Marginal PDF's from joint or conditional PDF's

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = \int_{-\infty}^{\infty} f_{X|Y}(x) f_Y(y)\,dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = \int_{-\infty}^{\infty} f_{Y|X}(y) f_X(x)\,dx$$

Joint Probability Density

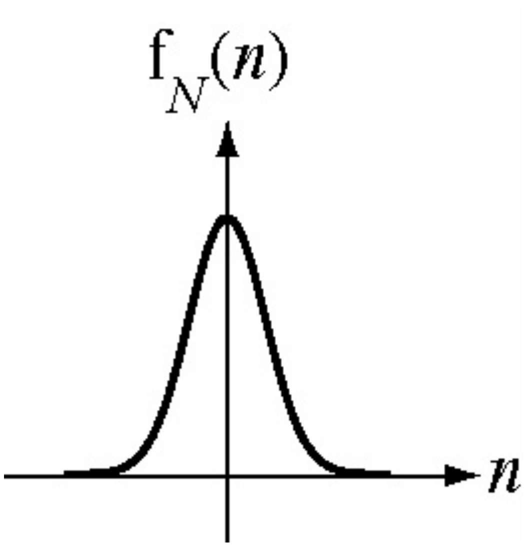
Example:

Let a message X with a known PDF be corrupted by additive noise N also with known pdf and received as $Y = X + N$. Then the best estimate that can be made of the message X is the value at the peak of the conditional PDF,

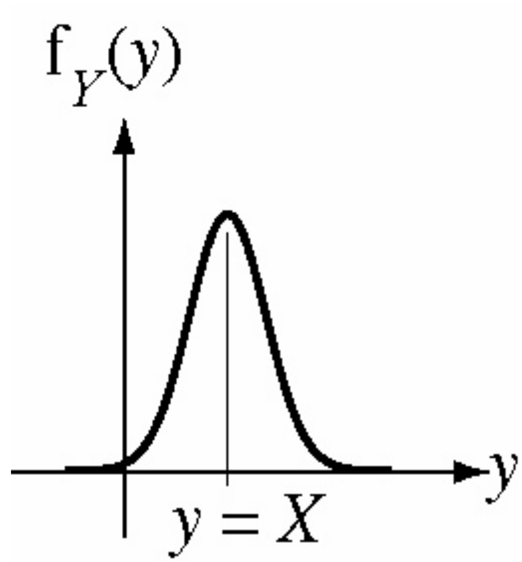
$$\mathrm{f}_{X|Y}(x) = \frac{\mathrm{f}_{Y|X}(y)\,\mathrm{f}_X(x)}{\mathrm{f}_Y(y)}$$

Joint Probability Density

Let N have the PDF,



Then, for any known value of X, the PDF of Y would be



Therefore if the PDF of N is $f_N(n)$, the conditional PDF of Y given X is $f_N(y - X)$

Joint Probability Density

Using Bayes' theorem,

$$\mathrm{f}_{X|Y}(x) = \frac{\mathrm{f}_{Y|X}(y)\mathrm{f}_X(x)}{\mathrm{f}_Y(y)} = \frac{\mathrm{f}_N(y-x)\mathrm{f}_X(x)}{\mathrm{f}_Y(y)}$$

$$= \frac{\mathrm{f}_N(y-x)\mathrm{f}_X(x)}{\int_{-\infty}^{\infty} \mathrm{f}_{Y|X}(y)\mathrm{f}_X(x)\,dx} = \frac{\mathrm{f}_N(y-x)\mathrm{f}_X(x)}{\int_{-\infty}^{\infty} \mathrm{f}_N(y-x)\mathrm{f}_X(x)\,dx}$$

Now the conditional PDF of X given Y can be computed.

Joint Probability Density

To make the example concrete let

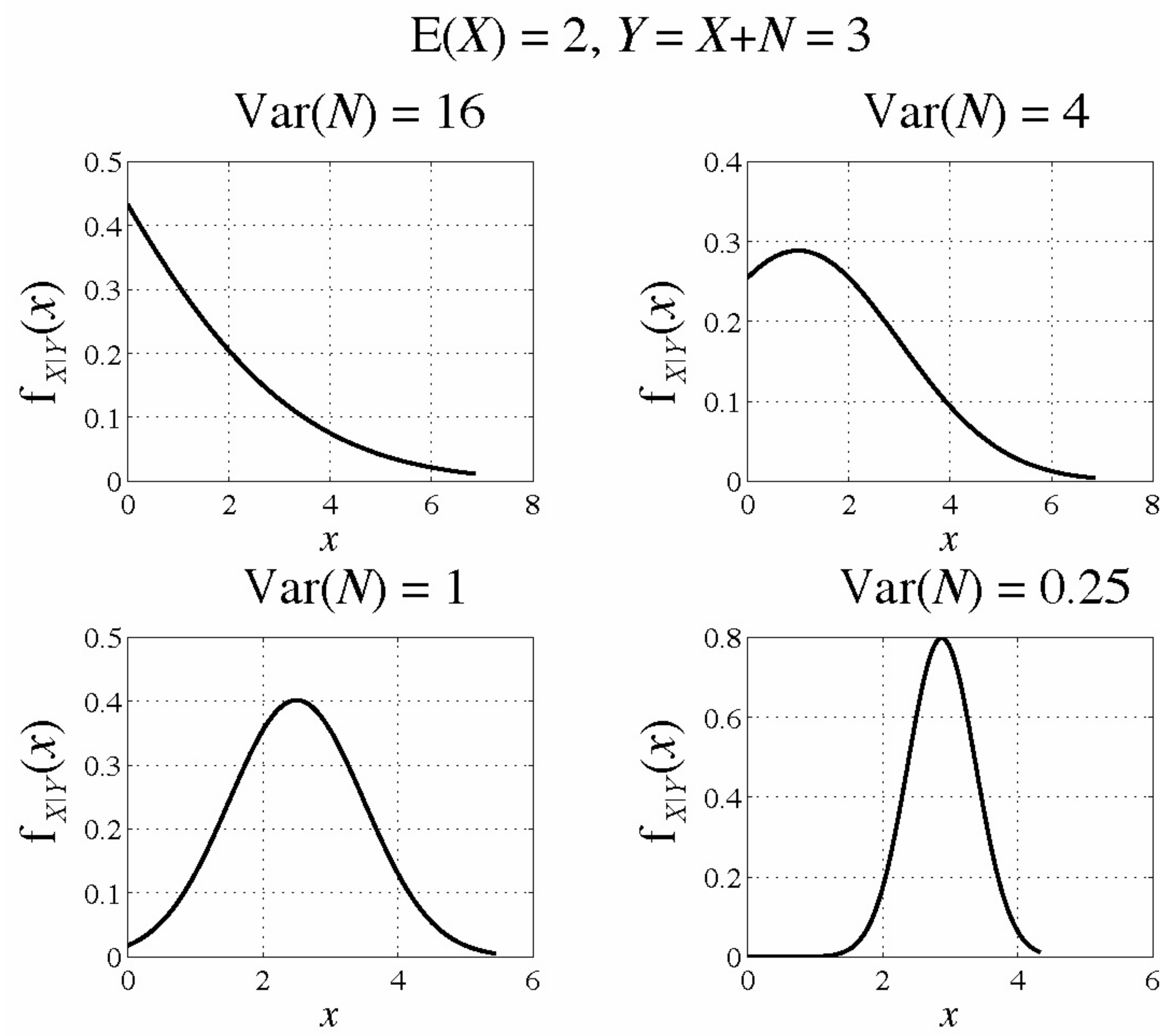
$$\mathrm{f}_X(x) = \frac{e^{-x/\mathrm{E}(X)}}{\mathrm{E}(X)}\mathrm{u}(x) \qquad \mathrm{f}_N(n) = \frac{1}{\sigma_N\sqrt{2\pi}}e^{-n^2/2\sigma_N^2}$$

Then the conditional pdf of X given Y is found to be

$$\mathrm{f}_Y(y) = \frac{\exp\left[\frac{\sigma_N^2}{2\,\mathrm{E}^2(X)} - \frac{y}{\mathrm{E}(X)}\right]}{2\,\mathrm{E}(X)}\left[1 + \mathrm{erf}\left(\frac{y - \frac{\sigma_N^2}{\mathrm{E}(X)}}{\sqrt{2}\sigma_N}\right)\right]$$

where **erf** is the **error function**.

Joint Probability Density



Independent Random Variables

If two random variables X and Y are independent then

$$\mathrm{f}_{X|Y}(x) = \mathrm{f}_X(x) = \frac{\mathrm{f}_{XY}(x,y)}{\mathrm{f}_Y(y)} \quad \text{and} \quad \mathrm{f}_{Y|X}(y) = \mathrm{f}_Y(y) = \frac{\mathrm{f}_{XY}(x,y)}{\mathrm{f}_X(x)}$$

Therefore $\mathrm{f}_{XY}(x,y) = \mathrm{f}_X(x)\,\mathrm{f}_Y(y)$ and their correlation is the product of their expected values.

$$\mathrm{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,\mathrm{f}_{XY}(x,y)\,dx\,dy = \int_{-\infty}^{\infty} y\,\mathrm{f}_Y(y)\,dy \int_{-\infty}^{\infty} x\,\mathrm{f}_X(x)\,dx = \mathrm{E}(X)\,\mathrm{E}(Y)$$

Independent Random Variables

Covariance

$$\sigma_{XY} \equiv \mathrm{E}\left(\left[X - \mathrm{E}(X)\right]\left[Y - \mathrm{E}(Y)\right]^{*}\right)$$

$$\sigma_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x - \mathrm{E}(X)\right)\left(y^{*} - \mathrm{E}(Y^{*})\right)\mathrm{f}_{XY}(x,y)\,dx\,dy$$

$$\sigma_{XY} = \mathrm{E}(XY^{*}) - \mathrm{E}(X)\mathrm{E}(Y^{*})$$

If X and Y are independent, $\sigma_{XY} = \mathrm{E}(X)\mathrm{E}(Y^{*}) - \mathrm{E}(X)\mathrm{E}(Y^{*}) = 0$

Independent Random Variables

Correlation Coefficient

$$\rho_{XY} = \mathrm{E}\left(\frac{X - \mathrm{E}(X)}{\sigma_X} \times \frac{Y^* - \mathrm{E}(Y^*)}{\sigma_Y} \right)$$

$$\rho_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{x - \mathrm{E}(X)}{\sigma_X} \right)\left(\frac{y^* - \mathrm{E}(Y^*)}{\sigma_Y} \right) \mathrm{f}_{XY}(x, y)\,dxdy$$

$$\rho_{XY} = \frac{\mathrm{E}(XY^*) - \mathrm{E}(X)\mathrm{E}(Y^*)}{\sigma_X \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

If X and Y are independent $\rho = 0$. If they are perfectly positively correlated $\rho = +1$ and if they are perfectly negatively correlated $\rho = -1$.

Independent Random Variables

If two random variables are independent, their covariance is zero. However, if two random variables have a zero covariance that does not mean they are necessarily independent.

Independence $\Rightarrow$ Zero Covariance

~~Zero Covariance $\Rightarrow$ Independence~~

Independent Random Variables

In the traditional jargon of random variable analysis, two “uncorrelated” random variables have a covariance of zero.

Unfortunately, this does not also imply that their correlation is zero. If their correlation is zero they are said to be **orthogonal**.

$$X \text{ and } Y \text{ are "Uncorrelated"} \Rightarrow \sigma_{XY} = 0$$

~~$X \text{ and } Y \text{ are "Uncorrelated"} \Rightarrow \mathrm{E}\left(XY \right) = 0$~~

Independent Random Variables

The variance of a sum of random variables X and Y is

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\sigma_{XY} = \sigma^2_X + \sigma^2_Y + 2\rho_{XY}\sigma_X\sigma_Y$$

If Z is a linear combination of random variables X_i

$$Z = a_0 + \sum_{i=1}^{N} a_i X_i$$

then $\mathrm{E}(Z) = a_0 + \sum_{i=1}^{N} a_i \mathrm{E}(X_i)$

$$\sigma^2_Z = \sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j \sigma_{X_i X_j} = \sum_{i=1}^{N} a_i^2 \sigma^2_{X_i} + \sum_{\substack{i=1\\ i\neq j}}^{N}\sum_{j=1}^{N} a_i a_j \sigma_{X_i X_j}$$

Independent Random Variables

If the X's are all independent of each other, the variance of the linear combination is a linear combination of the variances.

$$\sigma_Z^2 = \sum_{i=1}^{N} a_i^2 \sigma_{X_i}^2$$

If Z is simply the sum of the X's, and the X's are all independent of each other, then the variance of the sum is the sum of the variances.

$$\sigma_Z^2 = \sum_{i=1}^{N} \sigma_{X_i}^2$$

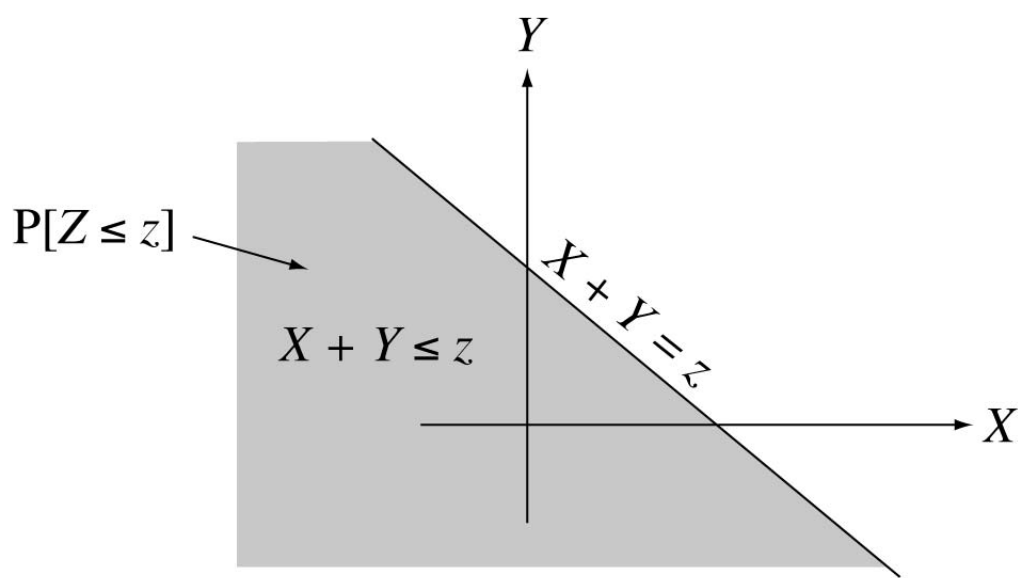
One Function of Two Random Variables

Let $Z = \mathrm{g}(X,Y)$. Find the pdf of Z.

$$\mathrm{F}_Z(z) = \mathrm{P}\left[Z \le z\right] = \mathrm{P}\left[\mathrm{g}(X,Y) \le z\right] = \mathrm{P}\left[(X,Y) \in R_Z\right]$$

where R_Z is the region in the XY plane where $\mathrm{g}(X,Y) \le z$

For example, let $Z = X + Y$



Probability Density of a Sum of Random Variables

Let $Z = X + Y$. Then for Z to be less than z, X must be less than $z - Y$. Therefore, the distribution function for Z is

$$\mathrm{F}_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} \mathrm{f}_{XY}(x, y)\,dx\,dy$$

If X and Y are independent, $\mathrm{F}_Z(z) = \int_{-\infty}^{\infty} \mathrm{f}_Y(y)\left(\int_{-\infty}^{z-y} \mathrm{f}_X(x)\,dx\right)dy$

and it can be shown that $\mathrm{f}_Z(z) = \int_{-\infty}^{\infty} \mathrm{f}_Y(y)\,\mathrm{f}_X(z-y)\,dy = \mathrm{f}_Y(z) * \mathrm{f}_X(z)$

Moment Generating Functions

The moment-generating function $\Phi_X(s)$ of a CV random variable

X is defined by $\Phi_X(s) = \mathrm{E}\left(e^{sX}\right) = \int_{-\infty}^{\infty} \mathrm{f}_X(x) e^{sx} dx.$

Relation to the Laplace transform $\rightarrow$ $\Phi_X(s) = \mathcal{L}\left[\mathrm{f}_X(x)\right]_{s \to -s}$

$$\frac{d}{ds}\left(\Phi_X(s)\right) = \int_{-\infty}^{\infty} \mathrm{f}_X(x) x e^{sx} dx$$

$$\left[\frac{d}{ds}\left(\Phi_X(s)\right)\right]_{s \to 0} = \int_{-\infty}^{\infty} x \, \mathrm{f}_X(x) dx = \mathrm{E}(X)$$

Relation to moments $\rightarrow$ $\mathrm{E}\left(X^n\right) = \left[\frac{d^n}{ds^n}\left(\Phi_X(s)\right)\right]_{s \to 0}$

Moment Generating Functions

The moment-generating function $\Phi_X(z)$ of a DV random variable X is defined by $\Phi_X(z) = \mathrm{E}\left(z^X\right) = \sum_{n=-\infty}^{\infty} \mathrm{P}\left[X = n\right] z^n = \sum_{n=-\infty}^{\infty} p_n z^n$.

Relation to the z transform $\rightarrow$ $\Phi_X(z) = \mathcal{Z}\left(\mathrm{P}_X(n)\right)_{z \to z^{-1}}$

$$\frac{d}{dz}\Phi_X(z) = \mathrm{E}\left(Xz^{X-1}\right) \qquad \frac{d^2}{dz^2}\Phi_X(z) = \mathrm{E}\left(X(X-1)z^{X-2}\right)$$

Relation to moments $\rightarrow$
$$\begin{cases} \left[\dfrac{d}{dz}\Phi_X(z)\right]_{z=1} = \mathrm{E}(X) \\ \left[\dfrac{d^2}{dz^2}\Phi_X(z)\right]_{z=1} = \mathrm{E}\left(X^2\right) - \mathrm{E}(X) \end{cases}$$

The Chebyshev Inequality

For any random variable X and any $\varepsilon > 0$,

$$\mathrm{P}\left[\left|X-\mu_X\right| \geq \varepsilon\right] = \int_{-\infty}^{-(\mu_X+\varepsilon)} \mathrm{f}_X(x)\,dx + \int_{\mu_X+\varepsilon}^{\infty} \mathrm{f}_X(x)\,dx = \int_{\left|X-\mu_X\right| \geq \varepsilon} \mathrm{f}_X(x)\,dx$$

Also

$$\sigma_X^2 = \int_{-\infty}^{\infty} \left(x-\mu_X\right)^2 \mathrm{f}_X(x)\,dx \geq \int_{\left|X-\mu_X\right| \geq \varepsilon} \left(x-\mu_X\right)^2 \mathrm{f}_X(x)\,dx \geq \varepsilon^2 \int_{\left|X-\mu_X\right| \geq \varepsilon} \mathrm{f}_X(x)\,dx$$

It then follows that $\mathrm{P}\left[\left|X-\mu_X\right| \geq \varepsilon\right] \leq \sigma_X^2 / \varepsilon^2$

This is known as the Chebyshev inequality. Using this we can put a bound on the probability of an event with knowledge only of the variance and no knowledge of the PMF or PDF.

The Markov Inequality

For any random variable X let $\mathrm{f}_X(x) = 0$ for all $X < 0$ and let ε be a postive constant. Then

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x\,\mathrm{f}_X(x)\,dx = \int_{0}^{\infty} x\,\mathrm{f}_X(x)\,dx \ge \int_{\varepsilon}^{\infty} x\,\mathrm{f}_X(x)\,dx \ge \varepsilon \int_{\varepsilon}^{\infty} \mathrm{f}_X(x)\,dx = \varepsilon\,\mathrm{P}[X \ge \varepsilon]$$

Therefore $\mathrm{P}[X \ge \varepsilon] \le \dfrac{\mathrm{E}(X)}{\varepsilon}$. This is known as the Markov inequality.

It allows us to bound the probability of certain events with knowledge only of the expected value of the random variable and no knowledge of the PMF or PDF except that it is zero for negative values.

The Weak Law of Large Numbers

Consider taking N independent values $\{X_1, X_2, \cdots, X_N\}$ from a random variable X in order to develop an understanding of the nature of X. They constitute a sampling of X. The sample mean is $\bar{X}_N = \frac{1}{N}\sum_{n=1}^{N} X_n$. The sample size is finite, so different sets of N values will yield different sample means. Thus $\bar{X}_N$ is itself a random variable and it is an estimator of the expected value of X, $\mathrm{E}(X)$. A good estimator has two important qualities. It is **unbiased** and **consistent**. Unbiased means $\mathrm{E}(\bar{X}_N) = \mathrm{E}(X)$. Consistent means that as N is increased the variance of the estimator is decreased.

The Weak Law of Large Numbers

Using the Chebyshev inequality we can put a bound on the probable deviation of $\bar{X}_N$ from its expected value.

$$\mathrm{P}\left[\left|\bar{X} - \mathrm{E}\left(\bar{X}_N\right)\right| \geq \varepsilon\right] \leq \frac{\sigma^2_{\bar{X}_N}}{\varepsilon^2} = \frac{\sigma^2_X}{N\varepsilon^2} \quad , \quad \varepsilon > 0$$

This implies that

$$\mathrm{P}\left[\left|\bar{X}_N - \mathrm{E}\left(X\right)\right| < \varepsilon\right] \geq 1 - \frac{\sigma^2_X}{N\varepsilon^2} \quad , \quad \varepsilon > 0$$

The probability that $\bar{X}_N$ is within some small deviation from $\mathrm{E}\left(X\right)$ can be made as close to one as desired by making N large enough.

The Weak Law of Large Numbers

Now, in

$$\mathrm{P}\left[\left|\bar{X}_N - \mathrm{E}(X)\right| < \varepsilon\right] \geq 1 - \frac{\sigma_X^2}{N\varepsilon^2} \quad , \quad \varepsilon > 0$$

let N approach infinity.

$$\lim_{N\to\infty} \mathrm{P}\left[\left|\bar{X}_N - \mathrm{E}(X)\right| < \varepsilon\right] = 1 \quad , \quad \varepsilon > 0$$

The **Weak Law of Large Numbers** states that if $\{X_1, X_2, \cdots, X_N\}$ is a sequence of iid random variable values and $\mathrm{E}(X)$ is finite, then

$$\lim_{N\to\infty} \mathrm{P}\left[\left|\bar{X}_N - \mathrm{E}(X)\right| < \varepsilon\right] = 1 \quad , \quad \varepsilon > 0$$

This kind of convergence is called **convergence in probability**.

The Strong Law of Large Numbers

Now consider a sequence $\{X_1, X_2, \cdots\}$ of independent values of X and let X have an expected value $E(X)$ and a finite variance σ_X^2. Also consider a sequence of sample means $\{\bar{X}_1, \bar{X}_2, \cdots\}$ defined by $\bar{X}_N = \frac{1}{N}\sum_{n=1}^{N} X_n$. The **Strong Law of Large Numbers** says

$$P\left[\lim_{N\to\infty} \bar{X}_N = E(X)\right] = 1$$

This kind of convergence is called **almost sure convergence**.

The Laws of Large Numbers

The Weak Law of Large Numbers

$$\lim_{N\to\infty} \mathrm{P}\left[\left|\bar{X}_N - \mathrm{E}(X)\right| < \varepsilon\right] = 1 \ , \ \ \varepsilon > 0$$

and the Strong Law of Large Numbers

$$\mathrm{P}\left[\lim_{N\to\infty} \bar{X}_N = \mathrm{E}(X)\right] = 1$$

seem to be saying about the same thing. There is a subtle difference. It can be illustrated by the following example in which a sequence converges in probability but not almost surely.

The Laws of Large Numbers

Let $X_{nk} = \begin{cases} 1 \ , \ k/n \le \zeta < (k+1)/n \ , \ 0 \le k < n \ , \ n = 1,2,3,\cdots \\ 0 \ , \ \text{otherwise} \end{cases}$

and let ζ be uniformly distributed between 0 and 1. As n increases from one we get this "triangular" sequence of X's.

$$\begin{matrix} X_{10} & & \\ X_{20} & X_{21} & \\ X_{30} & X_{31} & X_{32} \\ \vdots & & \end{matrix}$$

Now let $Y_{n(n-1)/2+k+1} = X_{nk}$ meaning that $Y = \{X_{10}, X_{20}, X_{21}, X_{30}, X_{31}, X_{32}, \cdots\}$.

X_{10} is one with probability one. X_{20} and X_{21} are each one with probability 1/2 and zero with probability 1/2. Generalizing we can say that X_{nk} is one with probability $1/n$ and zero with probability $1-1/n$.

The Laws of Large Numbers

$Y_{n(n-1)/2+k+1}$ is therefore one with probability $1/n$ and zero with probability $1-1/n$. For each n the probability that at least one of the n numbers in each length-n sequence is one is

$$\mathrm{P}\left[\text{at least one 1}\right]=1-\mathrm{P}\left[\text{no ones}\right]=1-\left(1-1/n\right)^{n}.$$

In the limit as n approaches infinity this probability approaches $1-1/e\cong 0.632$. So no matter how large n gets there is a non-zero probability that at least one 1 will occur in any length-n sequence. This proves that the sequence Y does not converge almost surely because there is always a non-zero probability that a length-n sequence will contain a 1 for any n.

The Laws of Large Numbers

The expected value $\mathrm{E}\left(X_{nk}\right)$ is

$$\mathrm{E}\left(X_{nk}\right) = \mathrm{P}\left[X_{nk} = 1\right] \times 1 + \mathrm{P}\left[X_{nk} = 0\right] \times 0 = 1/n$$

and is therefore independent of k and approaches zero as n approaches infinity. The expected value of X_{nk}^2 is

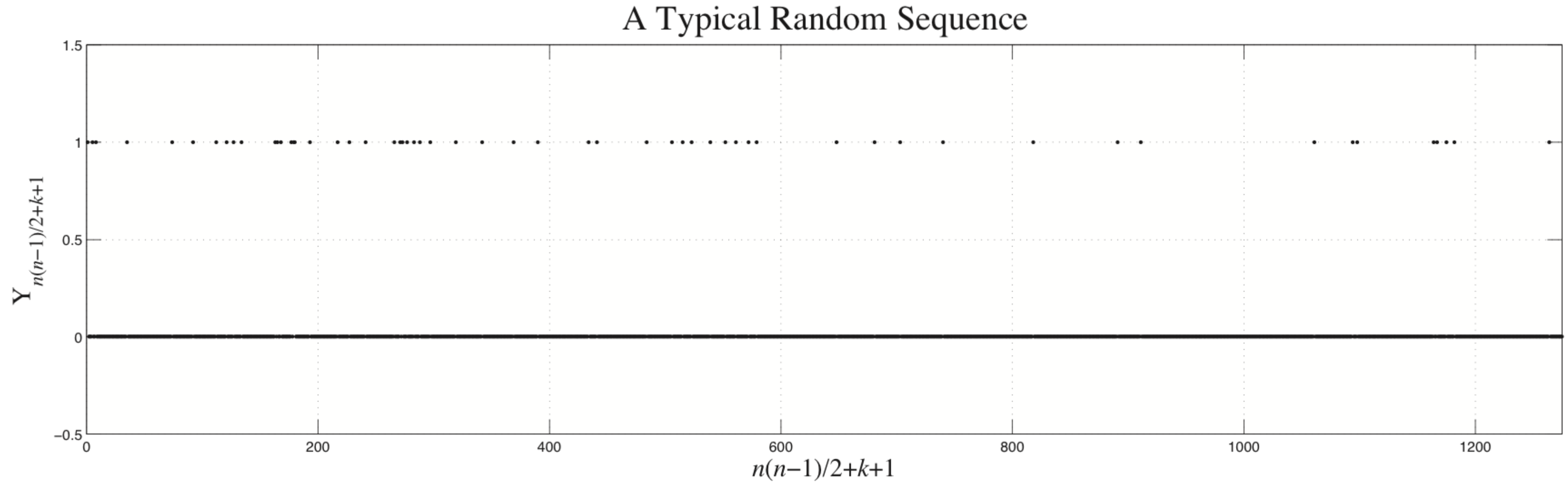
$$\mathrm{E}\left(X_{nk}^2\right) = \mathrm{P}\left[X_{nk} = 1\right] \times 1^2 + \mathrm{P}\left[X_{nk} = 0\right] \times 0^2 = \mathrm{E}\left(X_{nk}\right) = 1/n$$

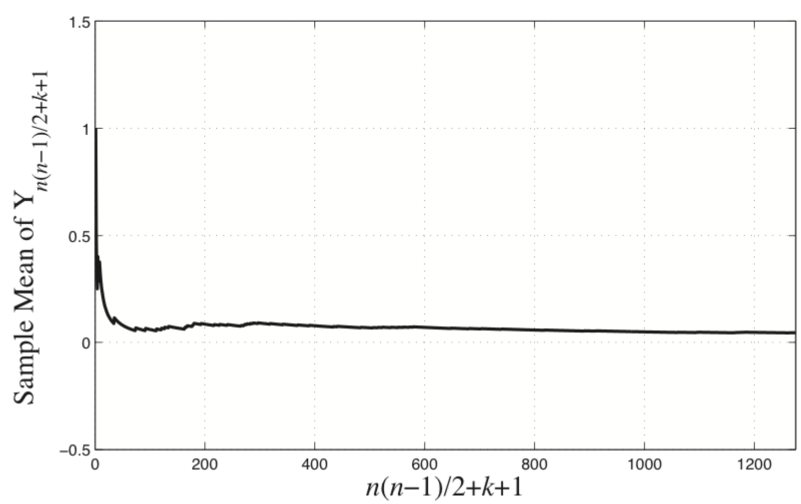
and the variance is X_{nk} is $\dfrac{n-1}{n^2}$. So the variance of Y approaches zero as n approaches infinity. Then according to the Chebyshev inequality

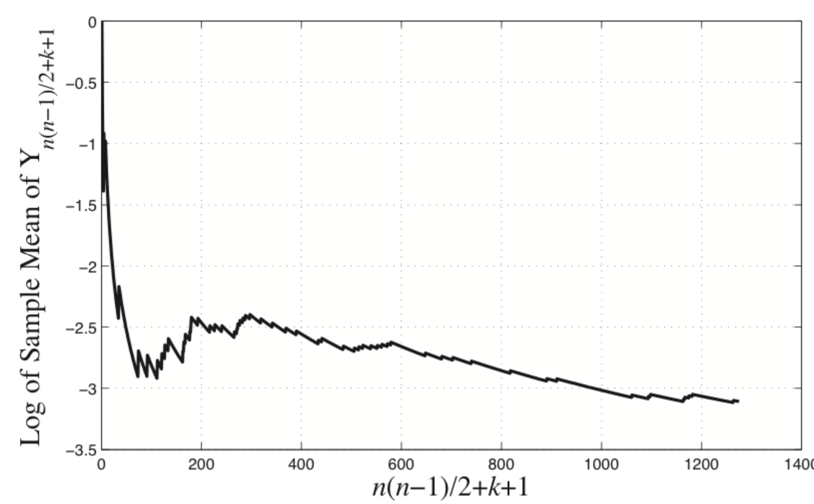
$$\mathrm{P}\left[\left|Y - \mu_Y\right| \geq \varepsilon\right] \leq \sigma_Y^2 / \varepsilon^2 = \frac{n-1}{n^2 \varepsilon^2}$$

implying that as n approaches infinity the variation of Y gets steadily smaller and that says that Y converges in probability to zero.

The Laws of Large Numbers

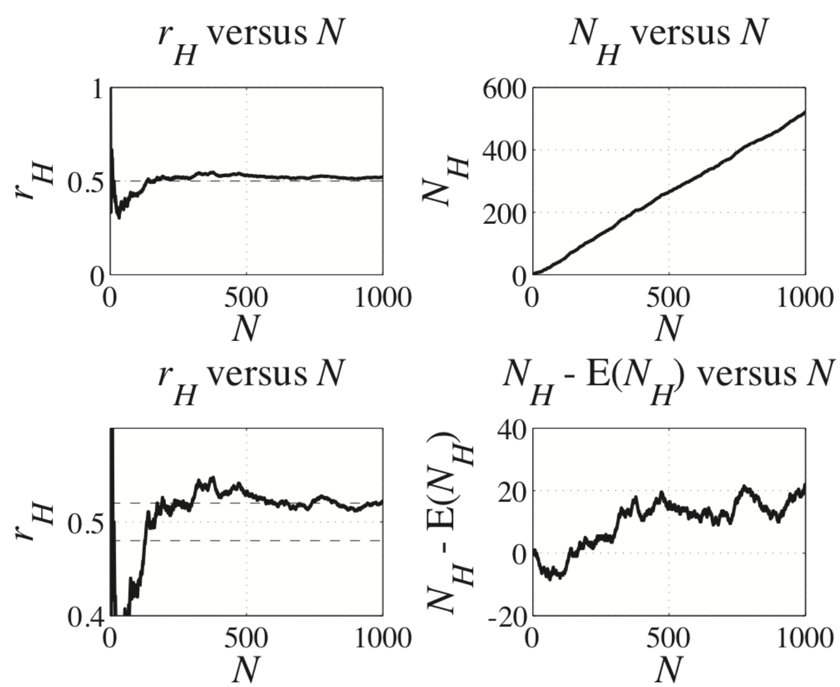






The Laws of Large Numbers

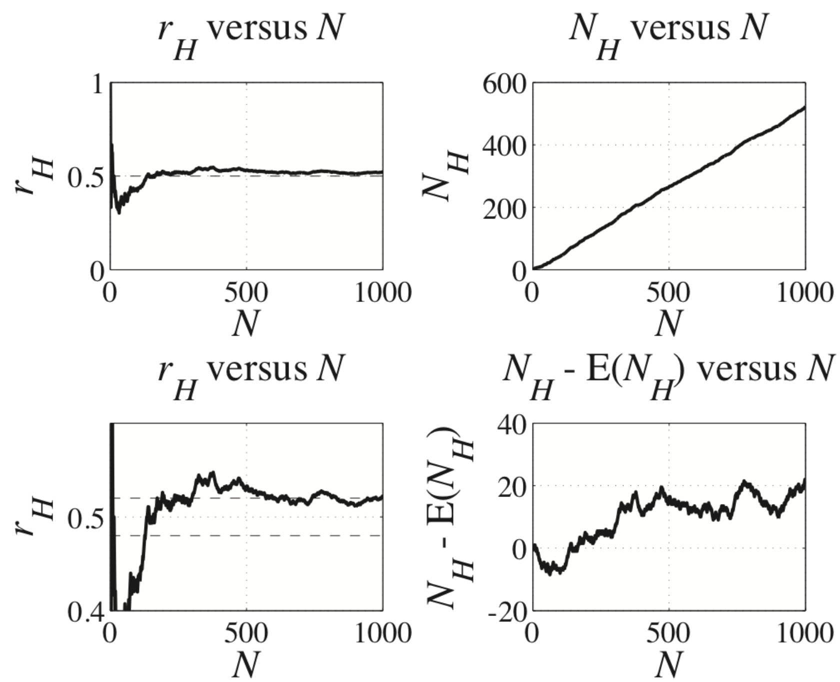
Consider an experiment in which we toss a fair coin and assign the value 1 to a head and the value 0 to a tail. Let N_H be the number of heads, let N be the number of coin tosses, let r_H be N_H / N and let X be the random variable indicating a head or tail. Then $N_H = \sum_{n=1}^{N} X_n$, $\mathrm{E}(N_H) = N/2$ and $\mathrm{E}(r_H) = 1/2$.



The Laws of Large Numbers

$\sigma^2_{r_H} = \sigma^2_X / N \Rightarrow \sigma_{r_H} = \sigma_X / \sqrt{N}$ Therefore $r_H - 1/2$ generally approaches zero but not smoothly or monotonically.

$\sigma^2_{N_H} = N\sigma^2_X \Rightarrow \sigma_{N_H} = \sqrt{N}\sigma_X$. Therefore $N_H - \mathrm{E}(N_H)$ does not approach zero. So the variation of N_H <u>increases</u> with N.



Convergence of Sequences of Random Variables

We have already seen two types of convergence of sequences of random variables, almost sure convergence (in the Strong Law of Large Numbers) and convergence in probability (in the Weak Law of Large Numbers). Now we will explore other types of convergence.

Convergence of Sequences of Random Variables

Sure Convergence

A sequence of random variables $\{X_n(\zeta)\}$ converges **surely** to the random variable $X(\zeta)$ if the sequence of functions $X_n(\zeta)$ converges to the function $X(\zeta)$ as $n \to \infty$ for <u>all</u> ζ in S. Sure convergence requires that every possible sequence converges. Different sequences may converge to different limits but all must converge.

$$X_n(\zeta) \to X(\zeta) \text{ as } n \to \infty \text{ for all } \zeta \in S$$

Convergence of Sequences of Random Variables

Almost Sure Convergence

A sequence of random variables $\{X_n(\zeta)\}$ converges **almost surely** to the random variable $X(\zeta)$ if the sequence of functions $X_n(\zeta)$ converges to the function $X(\zeta)$ as $n \to \infty$ for all ζ in S, except possible on a set of probability zero.

$$P[\zeta : X_n(\zeta) \to X(\zeta) \text{ as } n \to \infty] = 1$$

This is the convergence in the Strong Law of Large Numbers.

Convergence of Sequences of Random Variables

Mean Square Convergence

The sequence of random variables $\{X_n(\zeta)\}$ converges in the **mean - square** sense to the random variable $X(\zeta)$ if

$$E\left[\left(X_n(\zeta) - X(\zeta)\right)^2\right] \to 0 \text{ as } n \to \infty$$

If the limiting random variable $X(\zeta)$ is not known we can use the Cauchy Criterion: The sequence of random variables $\{X_n(\zeta)\}$ converges in the **mean - square** sense to the random variable $X(\zeta)$ if and only if

$$E\left[\left(X_n(\zeta) - X_m(\zeta)\right)^2\right] \to 0 \text{ as } n \to \infty \text{ and } m \to \infty$$

Convergence of Sequences of Random Variables

Convergence in Probability

The sequence of random variables $\{X_n(\zeta)\}$ converges **in probability** to the random variable $X(\zeta)$ if, for any $\varepsilon > 0$

$$P\left[\left|X_n(\zeta) - X(\zeta)\right| > \varepsilon\right] \to 0 \text{ as } n \to \infty$$

This is the convergence in the Weak Law of Large Numbers.

Convergence of Sequences of Random Variables

Convergence in Distribution

The sequence of random variables $\{X_n\}$ with cumulative distribution functions $\{F_n(x)\}$ converges **in distribution** to the random variable X with cumulative distribution function $F(x)$ if

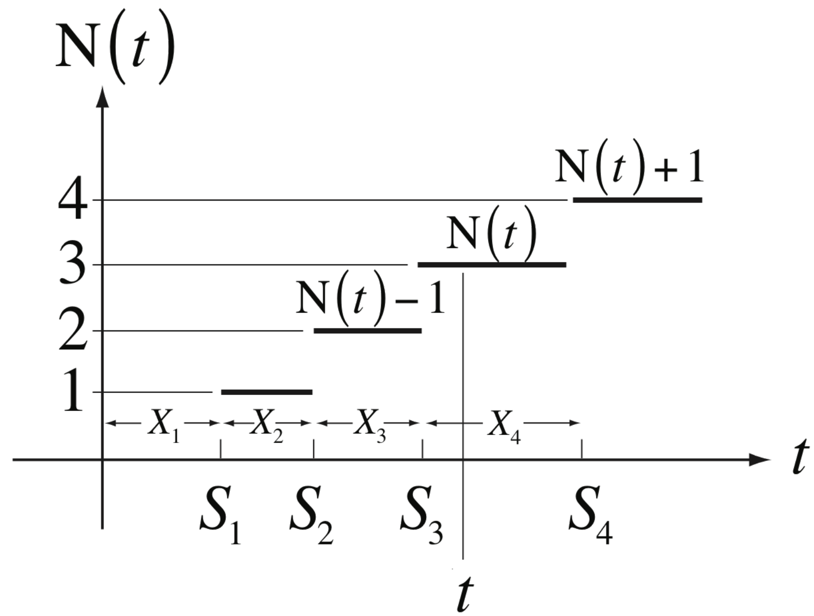
$$F_n(x) \rightarrow F(x) \text{ as } n \rightarrow \infty$$

for all x at which $F(x)$ is continuous. The Central Limit Theorem (coming soon) is an example of convergence in distribution.

Long-Term Arrival Rates

Suppose a system has a component that fails at time X_1, it is replaced and that component fails at time X_2, and so on. Let $\mathrm{N}(t)$ be the number of components that have failed at time t. $\mathrm{N}(t)$ is called a **renewal counting process**. Let X_j denote the lifetime of the jth component. Then the time when the nth component fails is $S_n = X_1 + X_2 + \cdots + X_n$ where we assume that the X_j are iid non-negative random variables with $0 \le \mathrm{E}(X) = \mathrm{E}(X_j) < \infty$.

We call the X_j's the **interarrival** or **cycle** times.



Long-Term Arrival Rates

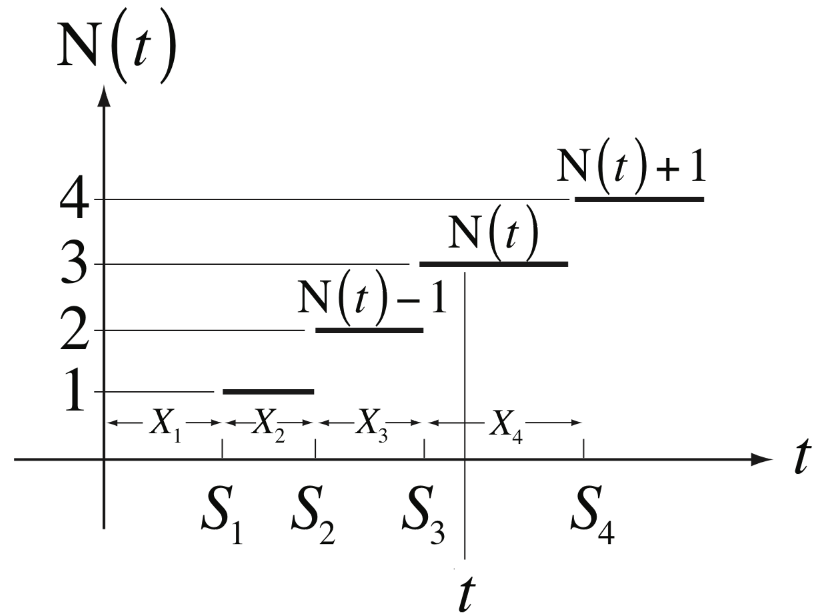
Since the average interarrival time is $\mathrm{E}(X)$ seconds per event one would expect intuitively that the average rate of arrivals is $1/\mathrm{E}(X)$ events per second.

$$S_{\mathrm{N}(t)} \leq t \leq S_{\mathrm{N}(t)+1}$$

Dividing through by $\mathrm{N}(t)$,

$$\frac{S_{\mathrm{N}(t)}}{\mathrm{N}(t)} \leq \frac{t}{\mathrm{N}(t)} \leq \frac{S_{\mathrm{N}(t)+1}}{\mathrm{N}(t)}$$

$\frac{S_{\mathrm{N}(t)}}{\mathrm{N}(t)}$ is the average interarrival time for the first $\mathrm{N}(t)$ arrivals.



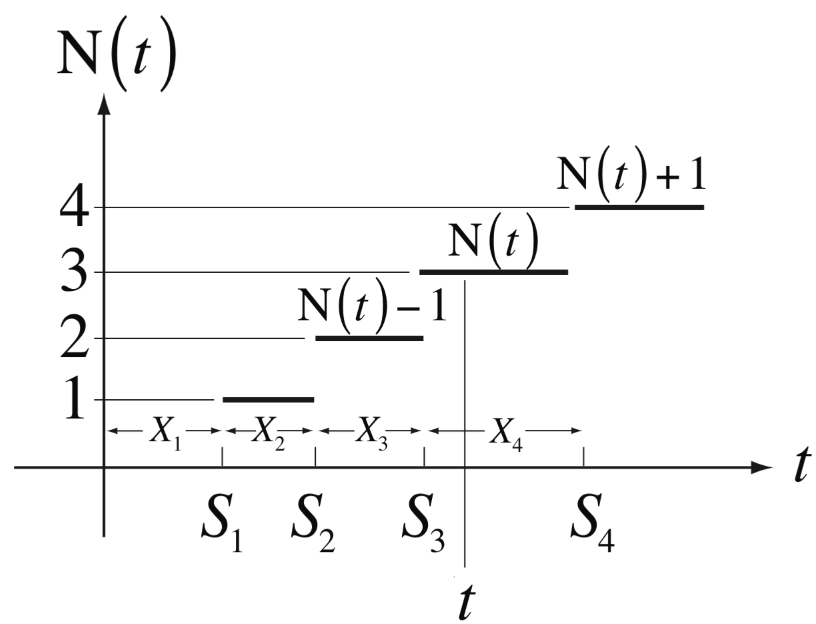
Long-Term Arrival Rates

$$\frac{S_{N(t)}}{N(t)} = \frac{1}{N(t)} \sum_{j=1}^{N(t)} X_j \quad \text{As } t \to \infty,\ N(t) \to \infty \text{ and } \frac{S_{N(t)}}{N(t)} \to E(X).$$

Similarly, $\dfrac{S_{N(t)+1}}{N(t)+1} \to E(X)$. So from $\dfrac{S_{N(t)}}{N(t)} \le \dfrac{t}{N(t)} \le \dfrac{S_{N(t)+1}}{N(t)}$

we can say $\lim\limits_{t\to\infty} \dfrac{t}{N(t)} = E(X)$ and

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{E(X)}.$$



Long-Term Time Averages

Suppose that events occur at random with iid interarrival times X_j and that a cost C_j is associated with each event. Let $\mathrm{C}_j(t)$ be the cost accumulated up to time t. Then $\mathrm{C}_j(t)=\sum_{j=1}^{\mathrm{N}(t)} C_j$. The average cost up to time t is $\frac{\mathrm{C}(t)}{t}=\frac{1}{t}\sum_{j=1}^{N(t)} C_j=\frac{\mathrm{N}(t)}{t}\frac{1}{\mathrm{N}(t)}\sum_{j=1}^{\mathrm{N}(t)} C_j$. In the limit $t\to\infty$,

$\frac{\mathrm{N}(t)}{t}\to\frac{1}{\mathrm{E}(X)}$ and $\frac{1}{\mathrm{N}(t)}\sum_{j=1}^{\mathrm{N}(t)} C_j\to\mathrm{E}(C)$. Therefore $\lim_{t\to\infty}\frac{\mathrm{C}(t)}{t}=\frac{\mathrm{E}(C)}{\mathrm{E}(X)}$.

The Central Limit Theorem

Let $Y_N = \sum_{n=1}^{N} X_n$ where the X_n's are an iid sequence of random variable values.

Let $Z_N = \frac{Y_N - N\,\mathrm{E}(X)}{\sigma_X \sqrt{N}} = \frac{\sum_{n=1}^{N}\left(X_n - \mathrm{E}(X)\right)}{\sigma_X \sqrt{N}}$.

$$\mathrm{E}(Z_N) = \mathrm{E}\left(\frac{\sum_{n=1}^{N}\left(X_n - \mathrm{E}(X)\right)}{\sigma_X \sqrt{N}}\right) = \frac{\sum_{n=1}^{N}\overbrace{\mathrm{E}\left(X_n - \mathrm{E}(X)\right)}^{=0}}{\sigma_X \sqrt{N}} = 0$$

The Central Limit Theorem

$$\sigma_{Z_N}^2 = \sum_{n=1}^{N}\left(\frac{1}{\sigma_X\sqrt{N}}\right)^2\sigma_X^2 = 1$$

The MGF of Z_N is $\Phi_{Z_N}(s) = \mathrm{E}\left(e^{sZ_N}\right) = \mathrm{E}\left(\exp\left(s\frac{\sum_{n=1}^{N}\left(X_n - \mathrm{E}(X)\right)}{\sigma_X\sqrt{N}}\right)\right)$.

$$\Phi_{Z_N}(s) = \mathrm{E}\left(\prod_{n-1}^{N}\exp\left(s\frac{\left(X_n - \mathrm{E}(X)\right)}{\sigma_X\sqrt{N}}\right)\right) = \prod_{n-1}^{N}\mathrm{E}\left(\exp\left(s\frac{\left(X_n - \mathrm{E}(X)\right)}{\sigma_X\sqrt{N}}\right)\right)$$

$$\Phi_{Z_N}(s) = \mathrm{E}^N\left(\exp\left(s\frac{\left(X - \mathrm{E}(X)\right)}{\sigma_X\sqrt{N}}\right)\right)$$

The Central Limit Theorem

We can expand the exponential function in an infinite series.

$$\Phi_{Z_N}(s) = \mathrm{E}^N\left(1 + s\frac{\left(X - \mathrm{E}(X)\right)}{\sigma_X\sqrt{N}} + s^2\frac{\left(X - \mathrm{E}(X)\right)^2}{2!\sigma_X^2 N} + s^3\frac{\left(X - \mathrm{E}(X)\right)^3}{3!\sigma_X^3 N\sqrt{N}} + \cdots \right)$$

$$\Phi_{Z_N}(s) = \left(1 + s\frac{\overbrace{\mathrm{E}\left(X - \mathrm{E}(X)\right)}^{=0}}{\sigma_X\sqrt{N}} + s^2\frac{\overbrace{\mathrm{E}\left(\left(X - \mathrm{E}(X)\right)^2\right)}^{=\sigma_X^2}}{2!\sigma_X^2 N} + s^3\frac{\mathrm{E}\left(\left(X - \mathrm{E}(X)\right)^3\right)}{3!\sigma_X^3 N\sqrt{N}} + \cdots \right)^N$$

$$\Phi_{Z_N}(s) = \left(1 + \frac{s^2}{2N} + s^3\frac{\mathrm{E}\left(\left(X - \mathrm{E}(X)\right)^3\right)}{3!\sigma_X^3 N\sqrt{N}} + \cdots \right)^N$$

The Central Limit Theorem

For large N we can neglect the higher-order terms. Then using

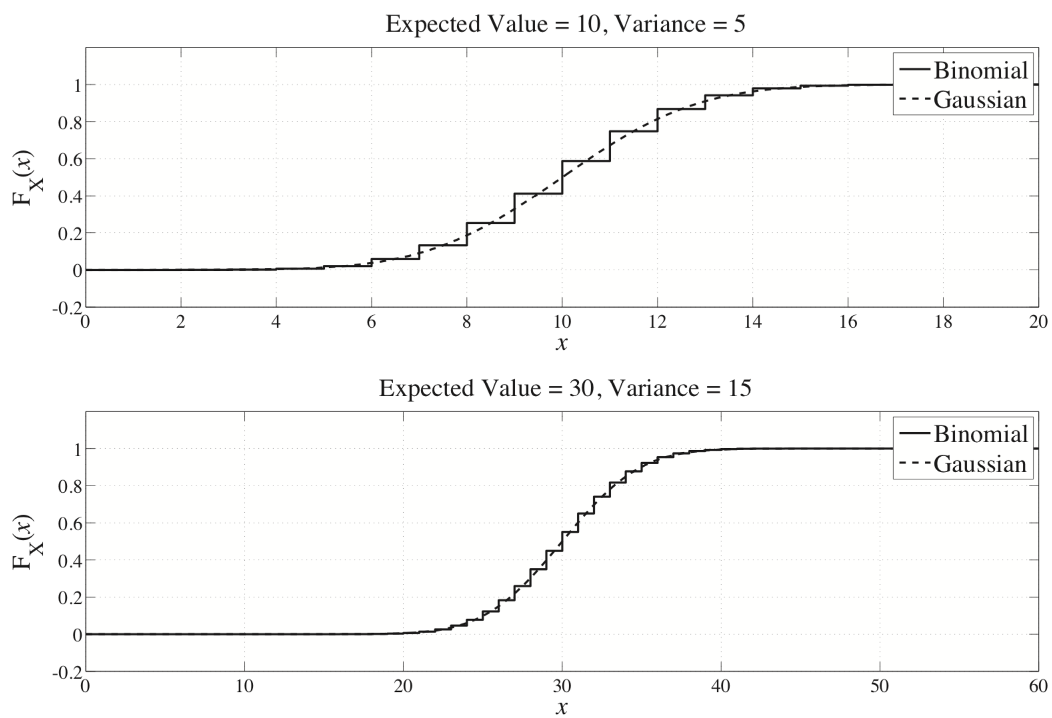
$$\lim_{m\to\infty}\left(1+\frac{z}{m}\right)^m = e^z \text{ we get}$$

$$\Phi_{Z_N}(s) = \lim_{N\to\infty}\left(1+\frac{s^2}{2N}\right)^N = e^{s^2/2} \Rightarrow \mathrm{f}_{Z_N}(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$$

Thus the PDF approaches a Gaussian shape, <u>with no assumptions</u> about the shapes of the PDF's of the X_n's. This is convergence in distribution.

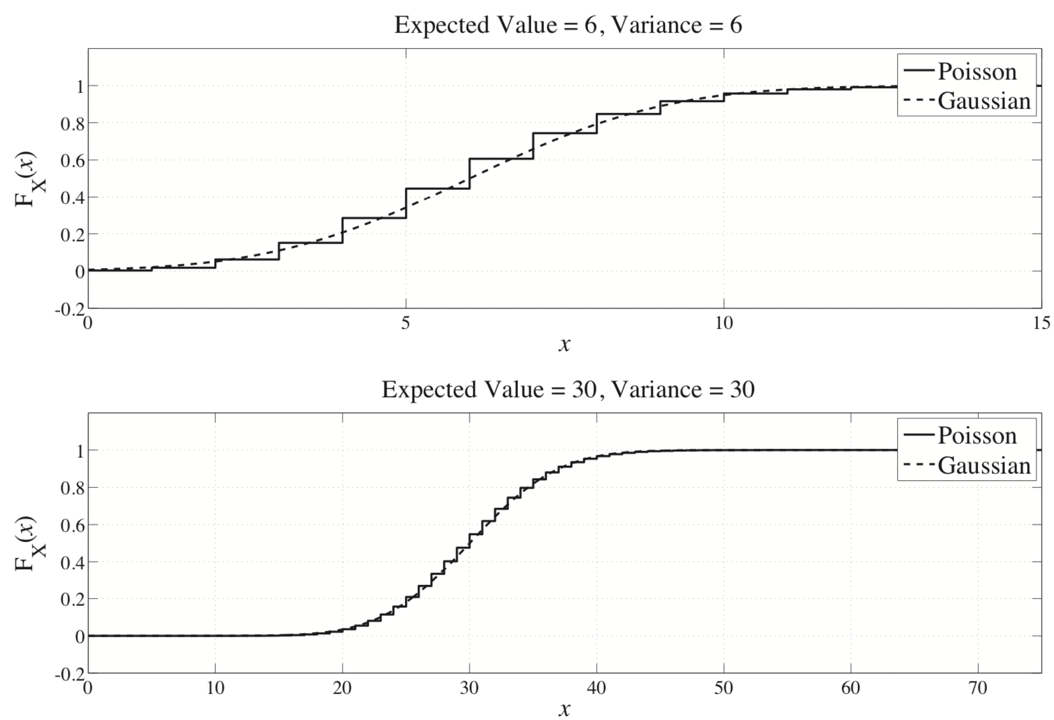
The Central Limit Theorem

Comparison of the distribution functions of two different Binomial random variables and Gaussian random variables with the same expected value and variance



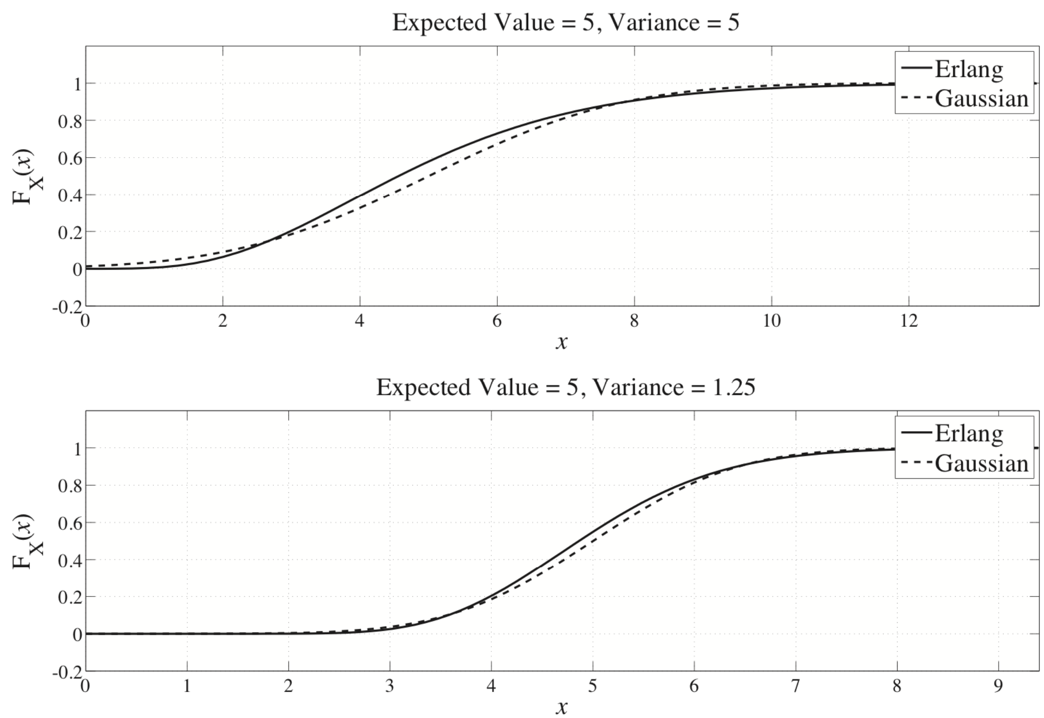
The Central Limit Theorem

Comparison of the distribution functions of two different Poisson random variables and Gaussian random variables with the same expected value and variance



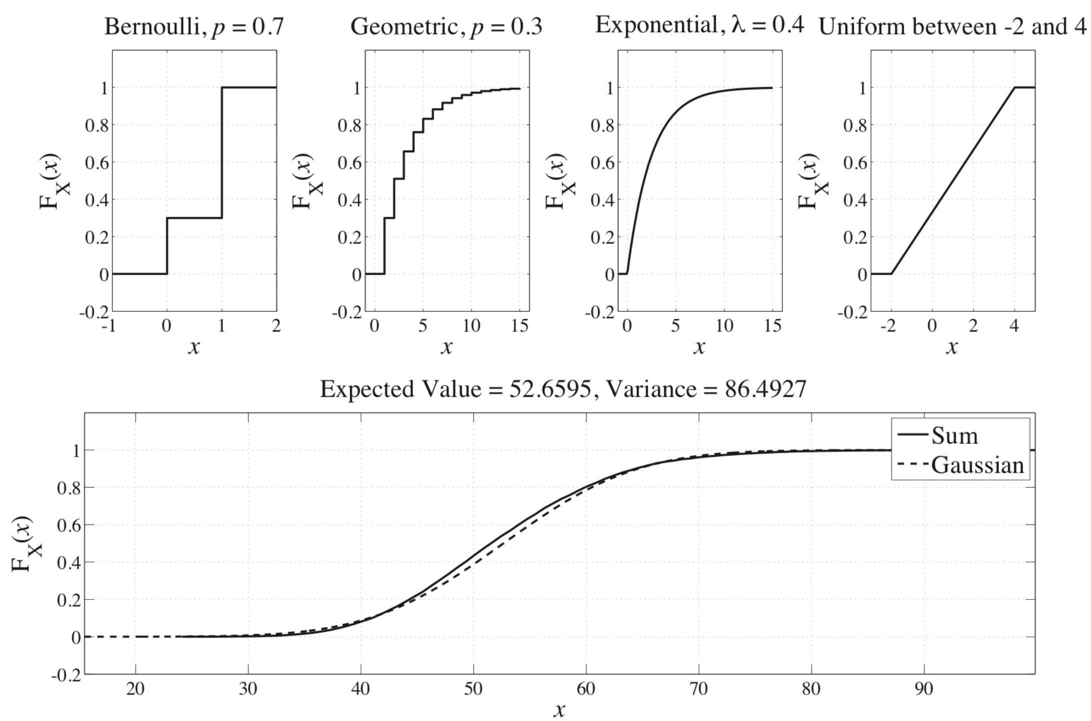
The Central Limit Theorem

Comparison of the distribution functions of two different Erlang random variables and Gaussian random variables with the same expected value and variance



The Central Limit Theorem

Comparison of the distribution functions of a sum of five independent random variables from each of four distributions and a Gaussian random variable with the same expected value and variance as that sum

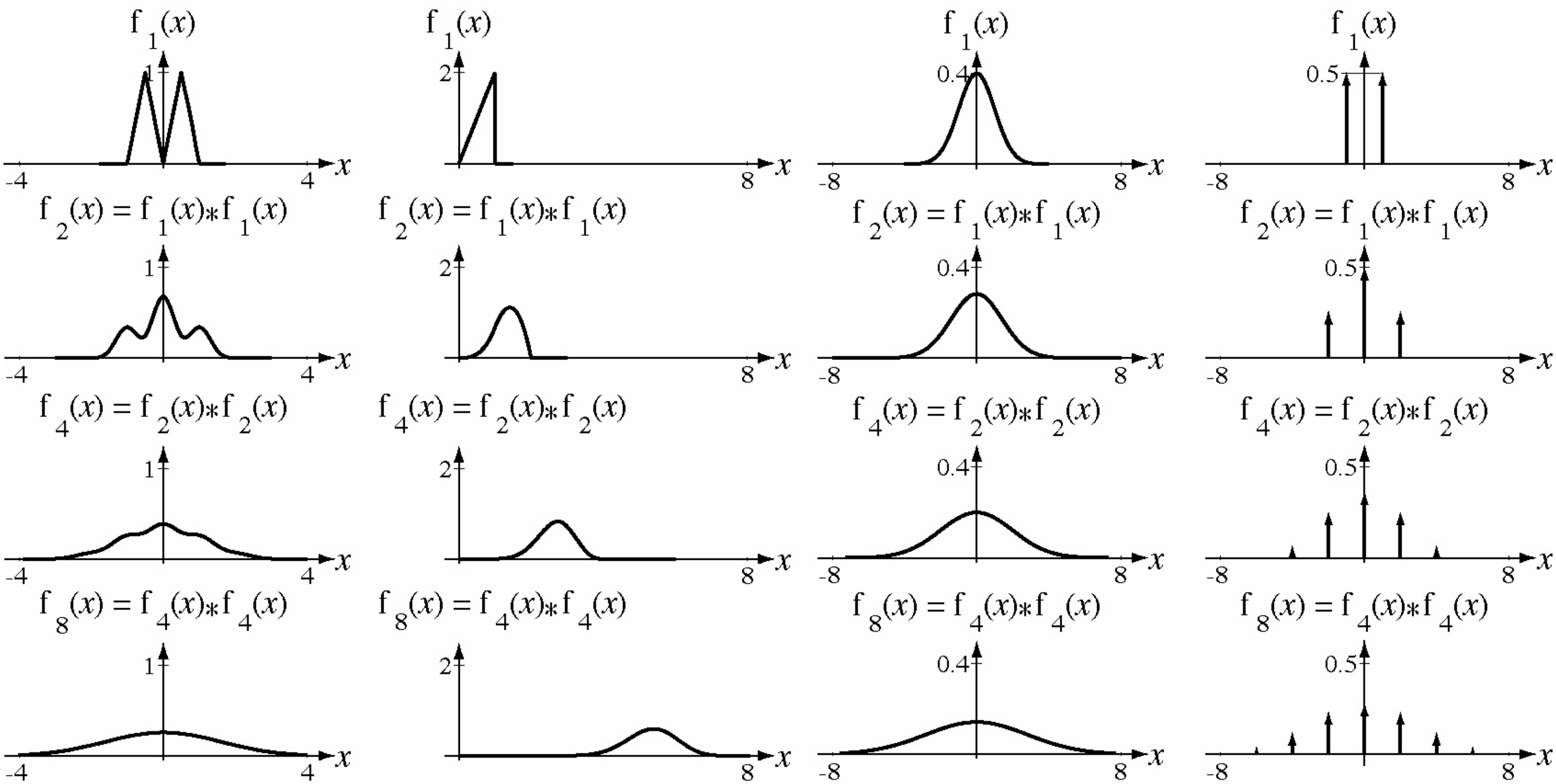


The Central Limit Theorem

The PDF of a sum of independent random variables is the convolution of their PDF's. This concept can be extended to any number of random variables. If $Z = \sum_{n=1}^{N} X_n$ then $f_Z(z) = f_{X_1}(z) * f_{X_2}(z) * f_{X_2}(z) * \cdots * f_{X_N}(z)$.

As the number of convolutions increases, the shape of the PDF of Z approaches the Gaussian shape.

The Central Limit Theorem

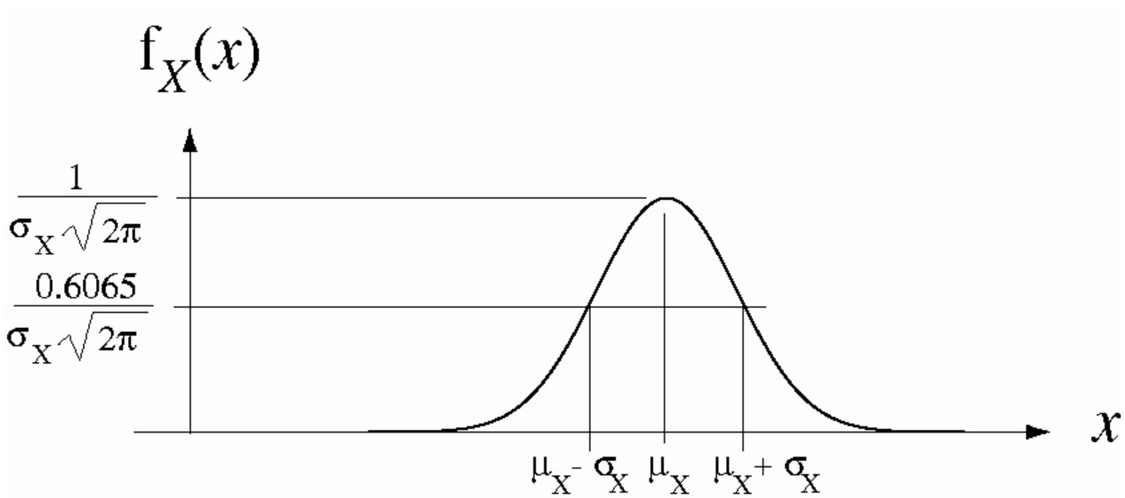


The Central Limit Theorem

The Gaussian pdf

$$\mathrm{f}_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} e^{-(x-\mu_X)^2/2\sigma_X^2}$$

$$\mu_X = \mathrm{E}(X) \text{ and } \sigma_X = \sqrt{\mathrm{E}\left(\left[X - \mathrm{E}(X)\right]^2\right)}$$



The Central Limit Theorem

The Gaussian PDF

Its maximum value occurs at the mean value of its argument.

It is symmetrical about the mean value.

The points of maximum absolute slope occur at one standard deviation above and below the mean.

Its maximum value is inversely proportional to its standard deviation.

The limit as the standard deviation approaches zero is a unit impulse.

$$\delta\left(x-\mu_x\right)=\lim_{\sigma_X \to 0} \frac{1}{\sigma_X \sqrt{2\pi}} e^{-\left(x-\mu_X\right)^2/2\sigma_X^2}$$

The Central Limit Theorem

The **normal** PDF is a Gaussian PDF with a mean of zero and a variance of one.

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

The central moments of the Gaussian PDF are

$$E\left(\left[X - E(X)\right]^n\right) = \begin{cases} 0 & , \ n \ \text{odd} \\ 1 \cdot 3 \cdot 5 \ldots (n-1)\sigma_X^n & , \ n \ \text{even} \end{cases}$$

The Central Limit Theorem

In computing probabilities from a Gaussian PDF it is necessary to evaluate integrals of the form, $\int_{x_1}^{x_2} \frac{dx}{\sigma_X \sqrt{2\pi}} e^{-(x-\mu_X)^2/2\sigma_X^2}$. Define a function

$\mathrm{G}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\lambda^2/2}\, d\lambda$. Then, using the change of variable $\lambda = \frac{x-\mu_X}{\sigma_X}$

we can convert the integral to $\int_{\frac{x_1-\mu_X}{\sigma_X}}^{\frac{x_2-\mu_X}{\sigma_X}} \frac{d\lambda}{\sqrt{2\pi}} e^{-\lambda^2/2}$ or $\mathrm{G}\left(\frac{x_2-\mu_X}{\sigma_X}\right) - \mathrm{G}\left(\frac{x_1-\mu_X}{\sigma_X}\right)$.

The G function is closely related to some other standard functions. For example the "error" function $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\lambda^2}\, d\lambda$ and $\mathrm{G}(x) = \frac{1}{2}\left(\mathrm{erf}\left(\sqrt{2}x\right)+1\right)$.

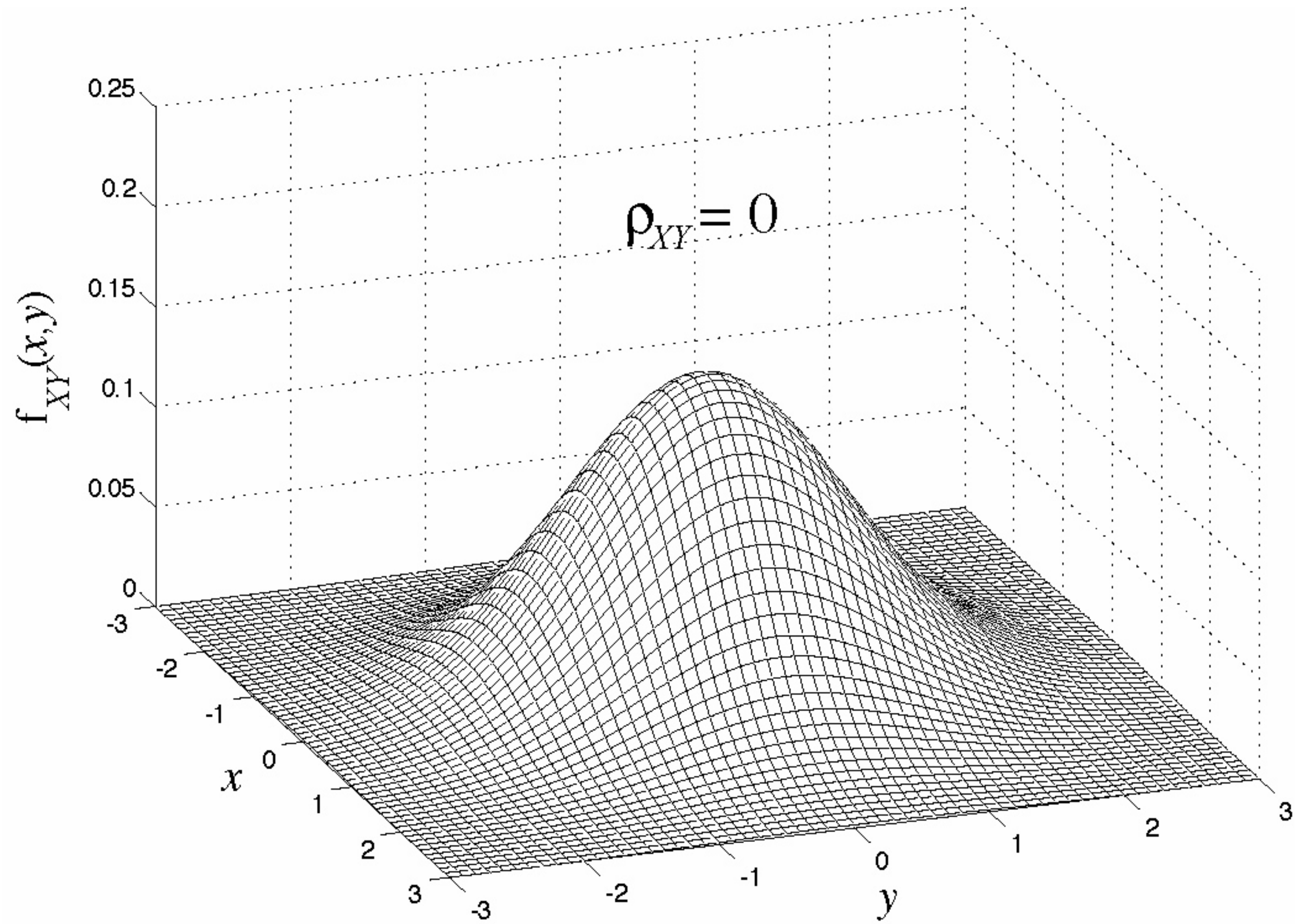
The Central Limit Theorem

Jointly Normal Random Variables

$$f_{XY}(x,y) = \frac{\exp\left[-\dfrac{\left(\dfrac{x-\mu_X}{\sigma_X}\right)^2 - \dfrac{2\rho_{XY}(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\dfrac{y-\mu_Y}{\sigma_Y}\right)^2}{2\left(1-\rho_{XY}^2\right)}\right]}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}}$$

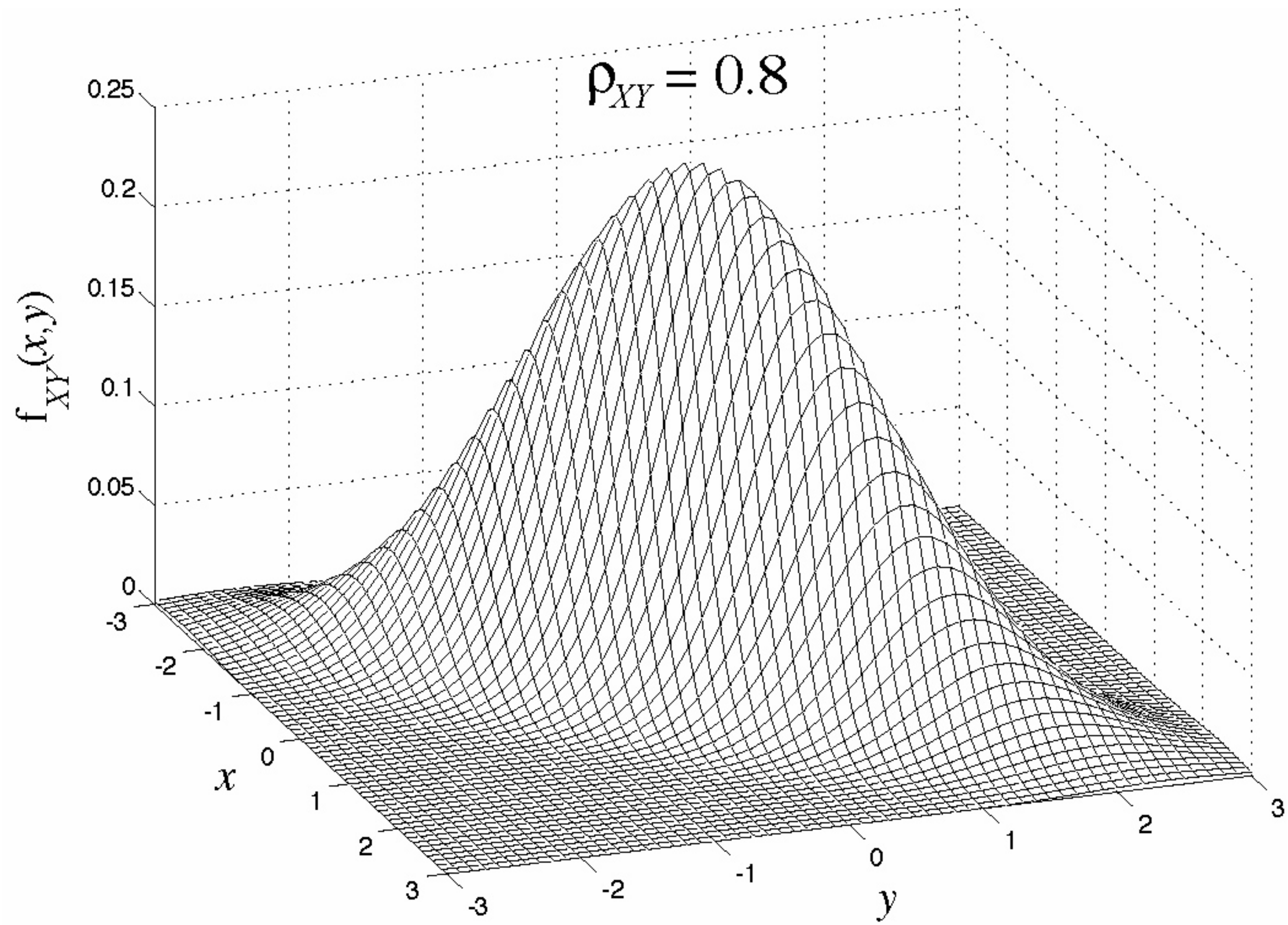
The Central Limit Theorem

Jointly Normal Random Variables



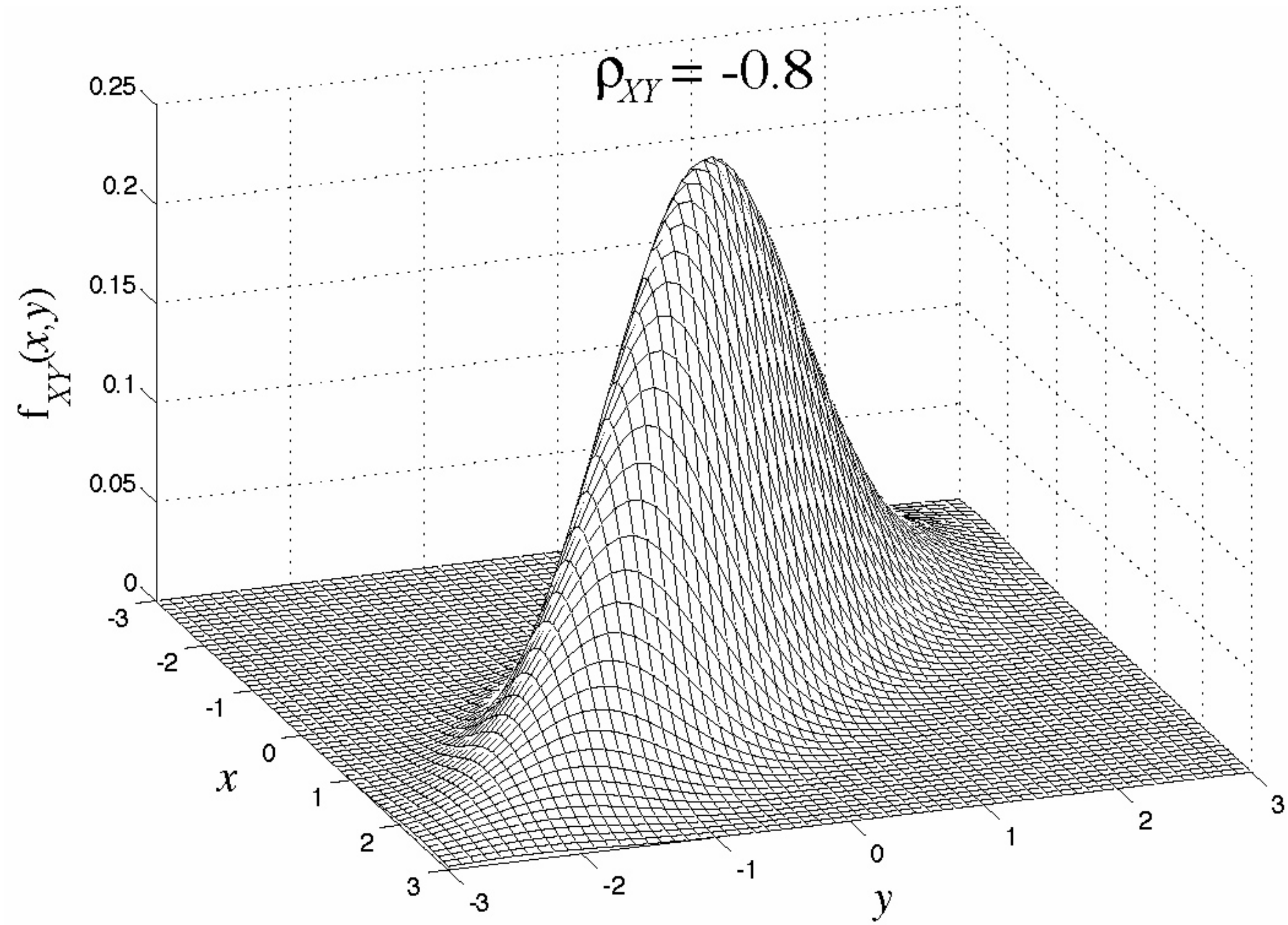
The Central Limit Theorem

Jointly Normal Random Variables



The Central Limit Theorem

Jointly Normal Random Variables



The Central Limit Theorem

Jointly Normal Random Variables

Any cross section of a bivariate Gaussian PDF at any value of x or y is a Gaussian. The marginal PDF's of X and Y can be found using

$$\mathrm{f}_X(x) = \int_{-\infty}^{\infty} \mathrm{f}_{XY}(x,y)\,dy$$

which turns out to be

$$\mathrm{f}_X(x) = \frac{e^{-(x-\mu_X)^2/2\sigma_X^2}}{\sigma_X\sqrt{2\pi}}$$

Similarly

$$\mathrm{f}_Y(y) = \frac{e^{-(y-\mu_Y)^2/2\sigma_Y^2}}{\sigma_Y\sqrt{2\pi}}$$

The Central Limit Theorem

Jointly Normal Random Variables

The conditional PDF of X given Y is

$$\mathrm{f}_{X|Y}\left(x\right)=\frac{\exp\left\{-\frac{\left[\left(x-\mu_X\right)-\left(\rho_{XY}\left(\sigma_X/\sigma_Y\right)\left(y-\mu_Y\right)\right)\right]^2}{2\sigma_X^2\left(1-\rho_{XY}^2\right)}\right\}}{\sqrt{2\pi}\sigma_X\sqrt{1-\rho_{XY}^2}}$$

The conditional PDF of Y given X is

$$\mathrm{f}_{Y|X}\left(y\right)=\frac{\exp\left\{-\frac{\left[\left(y-\mu_Y\right)-\left(\rho_{XY}\left(\sigma_Y/\sigma_X\right)\left(x-\mu_X\right)\right)\right]^2}{2\sigma_Y^2\left(1-\rho_{XY}^2\right)}\right\}}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho_{XY}^2}}$$

Transformations of Joint Probability Density Functions

If $W = \mathrm{g}(X,Y)$ and $Z = \mathrm{h}(X,Y)$ and both functions are invertible then it is possible to write $X = \mathrm{G}(W,Z)$ and $Y = \mathrm{H}(W,Z)$ and

$$\mathrm{P}\left[x < X \le x + \Delta x, y < Y \le y + \Delta y\right] = \mathrm{P}\left[w < W \le w + \Delta w, z < Z \le z + \Delta z\right]$$

$$\mathrm{f}_{XY}(x,y)\Delta x \Delta y \cong \mathrm{f}_{WZ}(w,z)\Delta w \Delta z$$

Transformations of Joint Probability Density Functions

$$\Delta x \Delta y = \left|J\right| \Delta w \Delta z \text{ where } \left|J\right| = \left\| \begin{matrix} \dfrac{\partial \mathrm{G}}{\partial w} & \dfrac{\partial \mathrm{G}}{\partial z} \\ \dfrac{\partial \mathrm{H}}{\partial w} & \dfrac{\partial \mathrm{H}}{\partial z} \end{matrix} \right\|$$

$$\mathrm{f}_{WZ}\left(w,z\right) = \left|J\right| \mathrm{f}_{XY}\left(x,y\right) = \left|J\right| \mathrm{f}_{XY}\left(\mathrm{G}\left(w,z\right), \mathrm{H}\left(w,z\right)\right)$$

Transformations of Joint Probability Density Functions

Let $R = \sqrt{X^2 + Y^2}$ and $\Theta = \tan^{-1}\left(\dfrac{Y}{X}\right)$, $-\pi < \Theta \le \pi$

where X and Y are independent and Gaussian, with zero mean and equal variances. Then

$$X = R\cos(\Theta) \quad \text{and} \quad Y = R\sin(\Theta)$$

$$|J| = \left\| \begin{matrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{matrix} \right\| = \left\| \begin{matrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{matrix} \right\| = r$$

Transformations of Joint Probability Density Functions

$$\mathrm{f}_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} e^{-x^2/2\sigma_X^2} \quad \text{and} \quad \mathrm{f}_Y(y) = \frac{1}{\sigma_Y\sqrt{2\pi}} e^{-y^2/2\sigma_Y^2}$$

Since X and Y are independent

$$\mathrm{f}_{XY}(x,y) = \frac{1}{2\pi\sigma^2} e^{-\left(x^2+y^2\right)/2\sigma^2} \qquad \sigma^2 = \sigma_X^2 = \sigma_Y^2$$

Applying the transformation formula

$$\mathrm{f}_{R\Theta}(r,\theta) = \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2}\,\mathrm{u}(r) \;,\; -\pi < \theta \le \pi$$

$$\mathrm{f}_{R\Theta}(r,\theta) = \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2}\,\mathrm{u}(r)\,\mathrm{rect}(\theta/2\pi)$$

Transformations of Joint Probability Density Functions

The radius R is distributed according to the Rayleigh PDF

$$\mathrm{f}_R(r) = \int_{-\pi}^{\pi} \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2}\,\mathrm{u}(r)\,d\theta = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}\,\mathrm{u}(r)$$

$$\mathrm{E}(R) = \sqrt{\frac{\pi}{2}}\sigma \quad \text{and} \quad \sigma_R^2 = 0.429\sigma^2$$

The angle is uniformly distributed

$$\mathrm{f}_\Theta(\theta) = \int_{-\infty}^{\infty} \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2}\,\mathrm{u}(r)\,dr = \frac{\mathrm{rect}(\theta/2\pi)}{2\pi} = \begin{cases} 1/2\pi & , \ -\pi < \theta \le \pi \\ 0 & , \ \text{otherwise} \end{cases}$$

Multivariate Probability Density

$$F_{X_1,X_2,\cdots,X_N}(x_1,x_2,\cdots,x_N) \equiv P\left[X_1 \le x_1 \cap X_2 \le x_2 \cap \cdots \cap X_N \le x_N\right]$$

$$0 \le F_{X_1,X_2,\cdots,X_N}(x_1,x_2,\cdots,x_N) \le 1 \quad , \quad -\infty < x_1 < \infty \quad , \quad \cdots \quad , \quad -\infty < x_N < \infty$$

$$F_{X_1,X_2,\cdots,X_N}(-\infty,\cdots,-\infty) = F_{X_1,X_2,\cdots,X_N}(-\infty,\cdots,x_k,\cdots,-\infty)$$

$$= F_{X_1,X_2,\cdots,X_N}(x_1,\cdots,-\infty,\cdots,x_N) = 0$$

$$F_{X_1,X_2,\cdots,X_N}(+\infty,\cdots,+\infty) = 1$$

$F_{X_1,X_2,\cdots,X_N}(x_1,x_2,\cdots,x_N)$ does not decrease if any number of $x's$ increase

$$F_{X_1,X_2,\cdots,X_N}(+\infty,\cdots,x_k,\cdots,+\infty) = F_{X_k}(x_k)$$

Multivariate Probability Density

$$\mathrm{f}_{X_1,X_2,\cdots,X_N}\left(x_1,x_2,\cdots,x_N\right)=\frac{\partial^N}{\partial x_1\partial x_2\cdots\partial x_N}\mathrm{F}_{X_1,X_2,\cdots,X_N}\left(x_1,x_2,\cdots,x_N\right)$$

$$\mathrm{f}_{X_1,X_2,\cdots,X_N}\left(x_1,x_2,\cdots,x_N\right)\geq 0\ \ ,\ -\infty<x_1<\infty\ \ ,\ \cdots\ \ ,\ -\infty<x_N<\infty$$

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{f}_{X_1,X_2,\cdots,X_N}\left(x_1,x_2,\cdots,x_N\right)dx_1dx_2\cdots dx_N=1$$

$$\mathrm{F}_{X_1,X_2,\cdots,X_N}\left(x_1,x_2,\cdots,x_N\right)=\int_{-\infty}^{x_N}\cdots\int_{-\infty}^{x_2}\int_{-\infty}^{x_1}\mathrm{f}_{X_1,X_2,\cdots,X_N}\left(\lambda_1,\lambda_2,\cdots,\lambda_N\right)d\lambda_1d\lambda_2\cdots d\lambda_N$$

$$\mathrm{f}_{X_k}\left(x_k\right)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{f}_{X_1,X_2,\cdots,X_N}\left(x_1,x_2,\cdots,x_{k-1},x_{k+1},\cdots,x_N\right)dx_1dx_2\cdots dx_{k-1}dx_{k+1}\cdots dx_N$$

$$\mathrm{P}\left[\left(X_1,X_2,\cdots,X_N\right)\in R\right]=\int\cdots\iint_R\mathrm{f}_{X_1,X_2,\cdots,X_N}\left(x_1,x_2,\cdots,x_N\right)dx_1dx_2\cdots dx_N$$

$$\mathrm{E}\left(\mathrm{g}\left(X_1,X_2,\cdots,X_N\right)\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{g}\left(x_1,x_2,\cdots,x_N\right)\mathrm{f}_{X_1,X_2,\cdots,X_N}\left(x_1,x_2,\cdots,x_N\right)dx_1dx_2\cdots dx_N$$

Other Important Probability Density Functions

In an ideal gas the three components of molecular velocity are all Gaussian with zero mean and equal variances of

$$\sigma_V^2 = \sigma_{V_X}^2 = \sigma_{V_Y}^2 = \sigma_{V_Z}^2 = kT / m$$

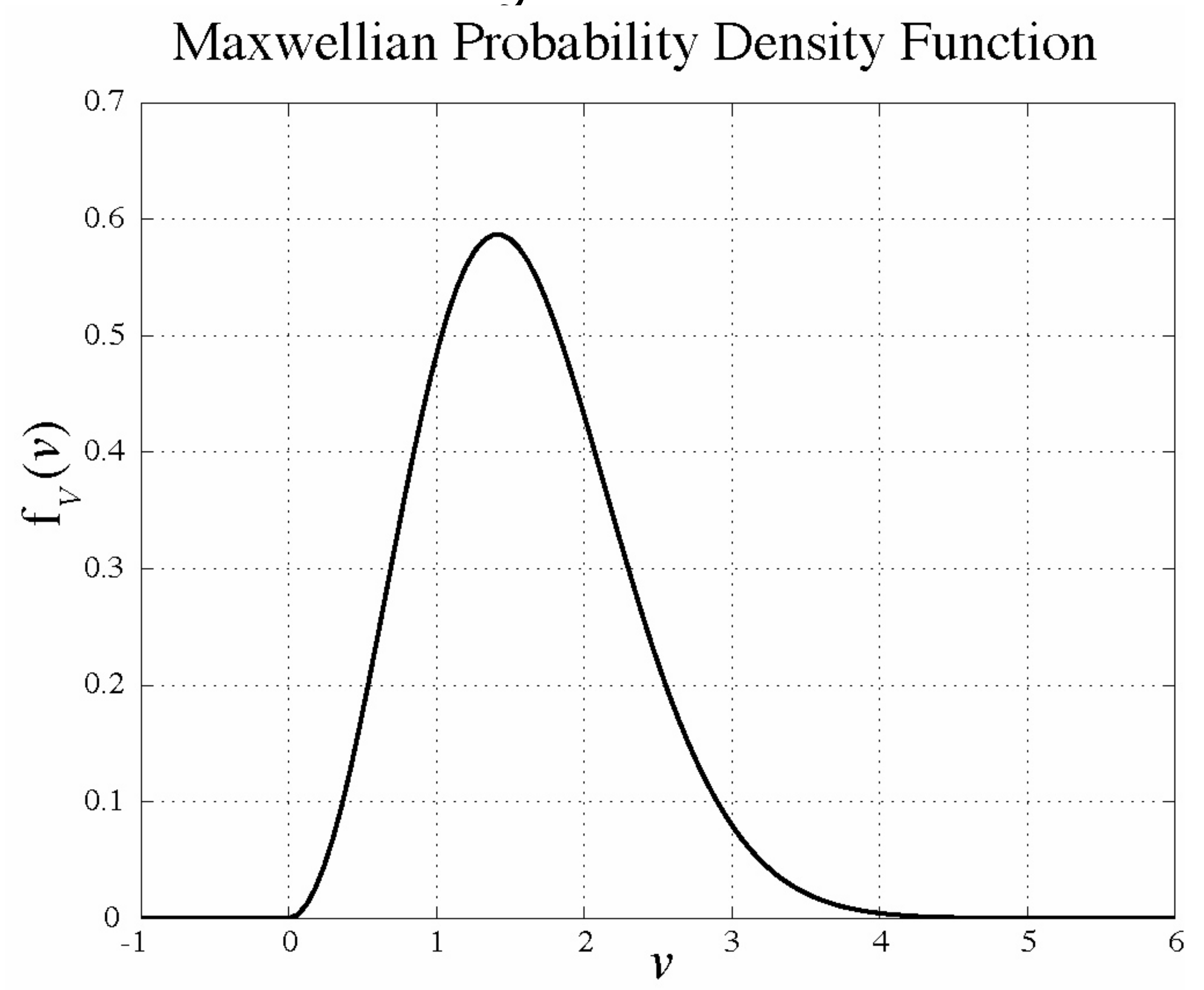
The speed of a molecule is

$$V = \sqrt{V_X^2 + V_Y^2 + V_Z^2}$$

and the PDF of the speed is called **Maxwellian** and is given by

$$\mathrm{f}_V\left(v\right) = \sqrt{2 / \pi}\,\frac{v^2}{\sigma_V^3}\,e^{-v^2/2\sigma_V^2}\,\mathrm{u}\left(v\right)$$

Other Important Probability Density Functions



Other Important Probability Density Functions

If $\chi^2 = Y_1^2 + Y_2^2 + Y_3^2 + \cdots + Y_N^2 = \sum_{n=1}^{N} Y_n^2$ and the random variables

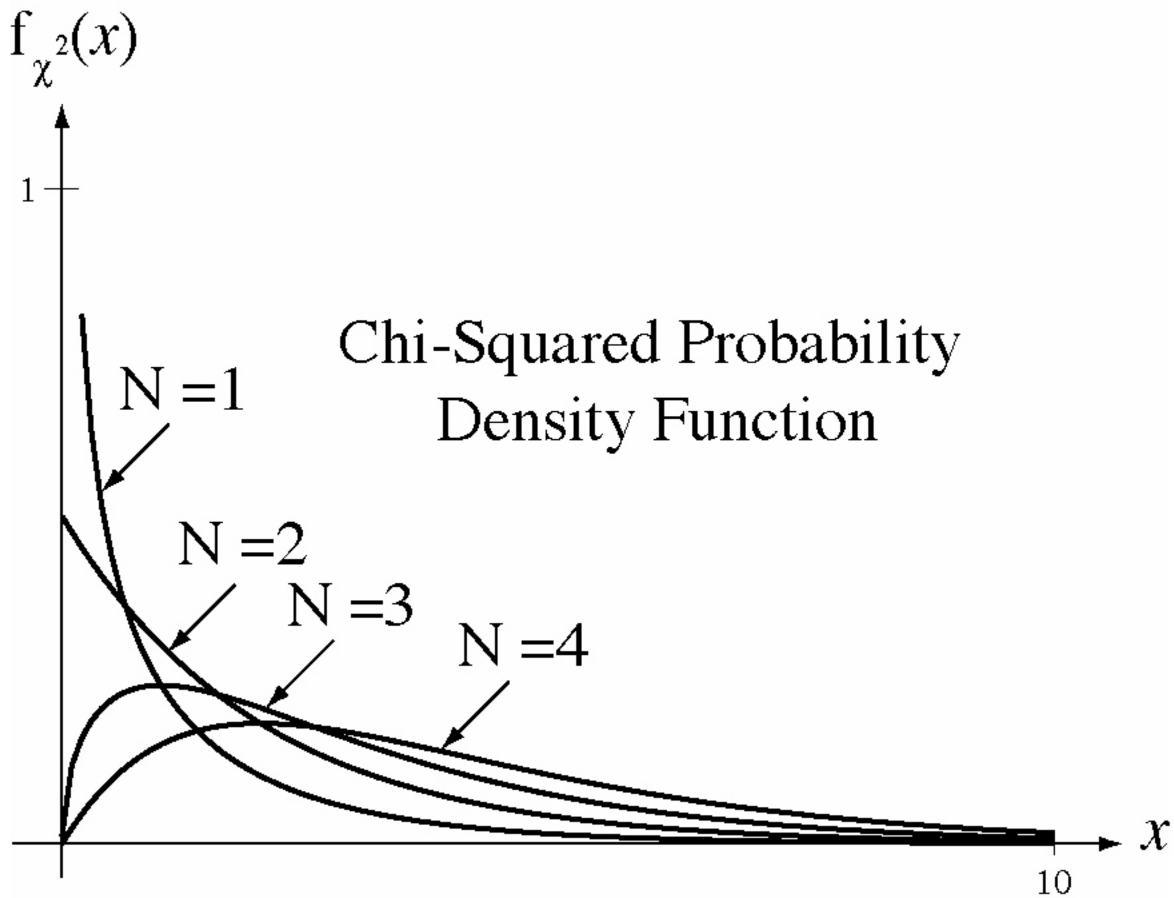
Y_n are all mutually independent and normally distributed then

$$\mathrm{f}_{\chi^2}(x) = \frac{x^{N/2-1}}{2^{N/2}\Gamma(N/2)} e^{-x/2}\,\mathrm{u}(x)$$

This is the **chi - squared** PDF.

$$\mathrm{E}(\chi^2) = N \qquad \sigma_{\chi^2}^2 = 2N$$

Other Important Probability Density Functions



Reliability

Reliability is defined by $\mathrm{R}(t) = \mathrm{P}[T > t]$ where T is the random variable representing the length of time after a system first begins operation that it fails.

$$\mathrm{F}_T(t) = \mathrm{P}[T \le t] = 1 - \mathrm{R}(t)$$

$$\frac{d}{dt}\big(\mathrm{R}(t)\big) = -\mathrm{f}_T(t)$$

Reliability

Probably the most commonly-used term in reliability analysis is **mean time to failure** (**MTTF**). MTTF is the expected value of T which is $\mathrm{E}(T) = \int_{-\infty}^{\infty} t\,\mathrm{f}_T(t)\,dt$. The conditional distribution function and PDF for the time to failure T given the condition $T > t_0$ are

$$\mathrm{F}_{T|T>t_0}(t) = \left\{ \begin{array}{ll} 0 & , \ t < t_0 \\ \dfrac{\mathrm{F}_T(t) - \mathrm{F}_T(t_0)}{1 - \mathrm{F}_T(t_0)} & , \ t \geq t_0 \end{array} \right\} = \frac{\mathrm{F}_T(t) - \mathrm{F}_T(t_0)}{\mathrm{R}(t_0)}\mathrm{u}(t - t_0)$$

$$\mathrm{f}_{T|T>t_0}(t) = \frac{\mathrm{f}_T(t)}{\mathrm{R}(t_0)}\mathrm{u}(t - t_0)$$

Reliability

A very common term in reliability analysis is failure rate which is defined by $\lambda(t)\,dt = \mathrm{P}\left[t < T \le t + dt\right] = \mathrm{f}_{T|T>t}(t)\,dt$. Failure rate is the probability per unit time that a system which has been operating properly up until time t will fail, as a function of t.

$$\lambda(t) = \frac{\mathrm{f}_T(t)}{\mathrm{R}(t)} = -\frac{\mathrm{R}'(t)}{\mathrm{R}(t)} \quad , \quad t \ge 0$$

$$\mathrm{R}'(t) + \lambda(t)\mathrm{R}(t) = 0 \quad , \quad t \ge 0$$

Reliability

The solution of $\mathrm{R}'(t)+\lambda(t)\mathrm{R}(t)=0\ ,\ t\geq 0$ is $\mathrm{R}(t)=e^{-\int_0^t \lambda(x)dx}\ ,\ t\geq 0.$

One of the simplest models for system failure used in reliability analysis is that the failure rate is a constant. Let that constant be K. Then

$\mathrm{R}(t)=e^{-\int_0^t K dx}=e^{-Kt}$ and $\mathrm{f}_T(t)=-\mathrm{R}'(t)=Ke^{-Kt}$ $\leftarrow$ Exponential PDF

MTTF is $1/K$.

Reliability

In some systems if any of the subsystems fails the overall system fails. If subsystem failure mechanisms are independent, the probability that the overall system is operating properly is the product of the probabilities that the subsystems are all operating properly. Let A_k be the event "subsystem k is operating properly" and let A_s be the event "the overall system is operating properly". Then, if there are N subsystems

$$\mathrm{P}\left[A_s\right] = \mathrm{P}\left[A_1\right]\mathrm{P}\left[A_2\right]\cdots\mathrm{P}\left[A_N\right] \text{ and } \mathrm{R}_s\left(t\right) = \mathrm{R}_1\left(t\right)\mathrm{R}_2\left(t\right)\cdots\mathrm{R}_N\left(t\right)$$

If the subsystems all have failure times with exponential PDF's then

$$\mathrm{R}_s\left(t\right) = e^{-t/\tau_1}e^{-t/\tau_2}\cdots e^{-t/\tau_N} = e^{-t\left(1/\tau_1+1/\tau_2+\cdots+1/\tau_N\right)} = e^{-t/\tau}$$

$$1/\tau = 1/\tau_1 + 1/\tau_2 + \cdots + 1/\tau_N$$

Reliability

In some systems the overall system fails only if all of the subsystems fail . If subsystem failure mechanisms are independent, the probability that the overall system is not operating properly is the product of the probabilities that the subsystems are all not operating properly. As before let A_k be the event "subsystem k is operating properly" and let A_s be the event "the overall system is operating properly". Then, if there are N subsystems

$$\mathrm{P}\left[\bar{A}_s\right] = \mathrm{P}\left[\bar{A}_1\right]\mathrm{P}\left[\bar{A}_2\right]\cdots\mathrm{P}\left[\bar{A}_N\right]$$

and $1-\mathrm{R}_s(t) = \left(1-\mathrm{R}_1(t)\right)\left(1-\mathrm{R}_2(t)\right)\cdots\left(1-\mathrm{R}_N(t)\right)$

If the subsystems all have failure times with exponential PDF's then

$$\mathrm{R}_s(t) = 1-\left(1-e^{-t/\tau_1}\right)\left(1-e^{-t/\tau_2}\right)\cdots\left(1-e^{-t/\tau_N}\right)$$

Reliability

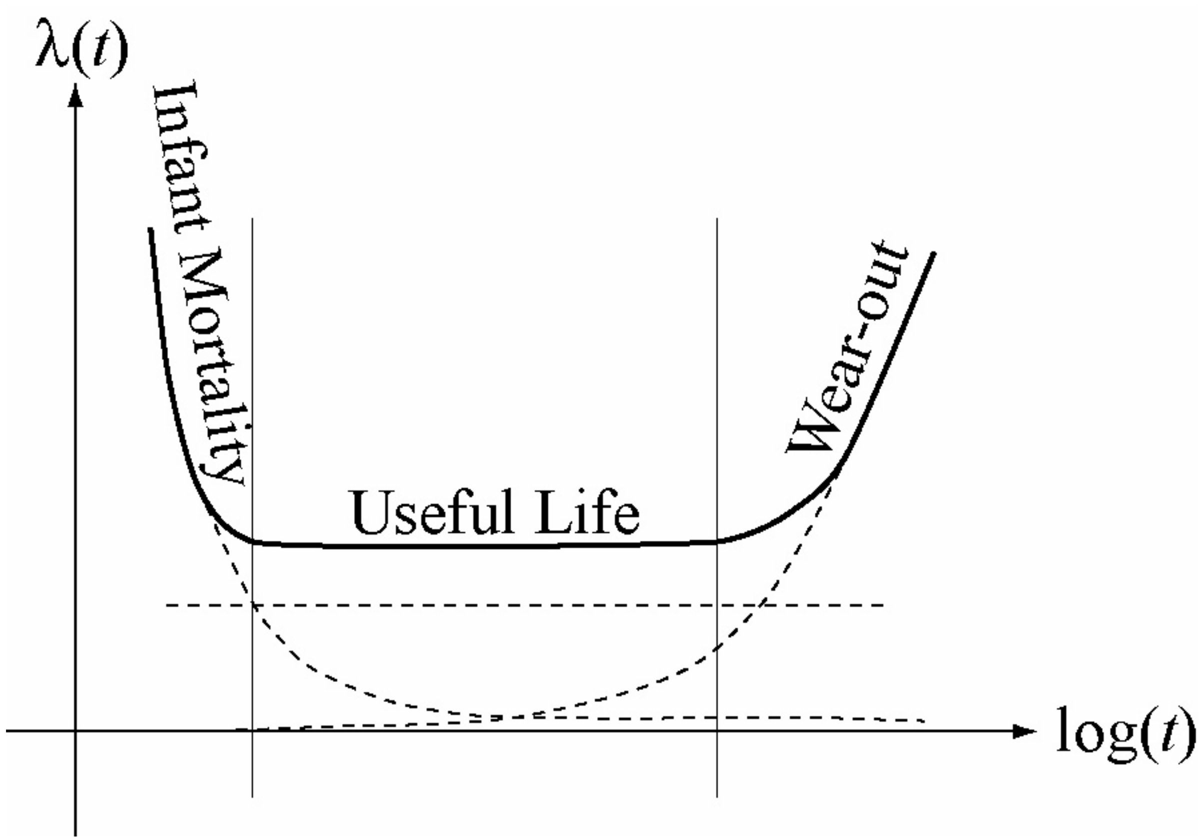
An exponential failure rate implies that whether a system has just begun operation or has been operating properly for a long time, the probability that it will fail in the next unit of time is the same. The expected value of the additional time to failure at any arbitrary time is a constant independent of past history,

$$\mathrm{E}\left(T \mid T > t_0\right) = t_0 + \mathrm{E}\left(T\right)$$

This model is fairly reasonable for a wide range of times but not for all times in all systems. Many real systems experience two additional types of failure that are not indicated by an exponential PDF of failure times, **infant mortality** and **wear - out**.

Reliability

The "Bathtub" Curve



Reliability

The two higher-failure-rate portions of the bathtub curve are often modeled by the **log - normal** distribution of failure times.

If a random variable X is Gaussian distributed its PDF is

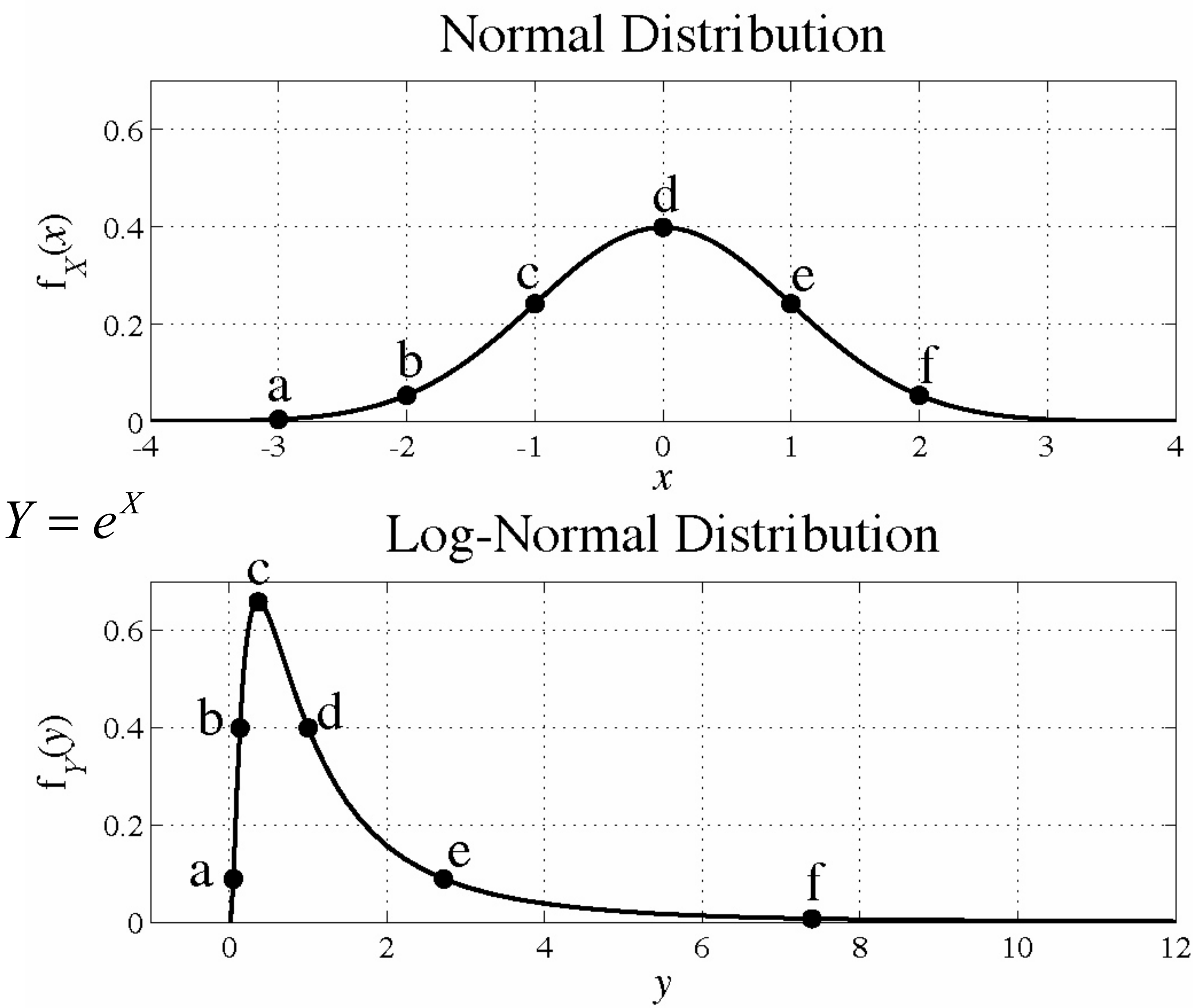
$$\mathrm{f}_X(x) = \frac{e^{-(x-\mu_X)^2/2\sigma_X^2}}{\sigma_X\sqrt{2\pi}}$$

If $Y = e^X$ then $dY/dX = e^X = Y, X = \ln(Y)$ and the PDF of Y is

$$\mathrm{f}_Y(y) = \frac{\mathrm{f}_X(\ln(y))}{|dy/dx|} = \frac{e^{-(\ln(y)-\mu_X)^2/2\sigma_X^2}}{y\sigma_X\sqrt{2\pi}}$$

Y is log-normal distributed $\mathrm{E}(Y) = e^{\mu_X+\sigma_X^2/2}$ and $\sigma_Y^2 = e^{2\mu_X+\sigma_X^2}\left(e^{\sigma_X^2}-1\right)$.

The Log-Normal Distribution



The Log-Normal Distribution

Another common application of the log-normal distribution is to model the pdf of a random variable X that is formed from the product of a large number N of independent random variables X_n.

$$X = \prod_{n=1}^{N} X_n$$

The logarithm of X is then

$$\log(X) = \sum_{n=1}^{N} \log(X_n)$$

Since $\log(X)$ is the sum of a large number of independent random variables its PDF tends to be Gaussian which implies that the PDF of X is log-normal in shape.