Principles of Communications Lecture 9: A Brief Review of Probability Theory

Chih-Wei Liu 劉志尉
National Chiao Tung University
cwliu@twins.ee.nctu.edu.tw

Outlines

- What is Probability
- Random Variables
- Distribution Functions
- Multiple Variables
- Statistical Averages
- Useful pdfs

Motivation

- Both messages (signals) and noises are random (stochastic) in nature.
- Some definitions:

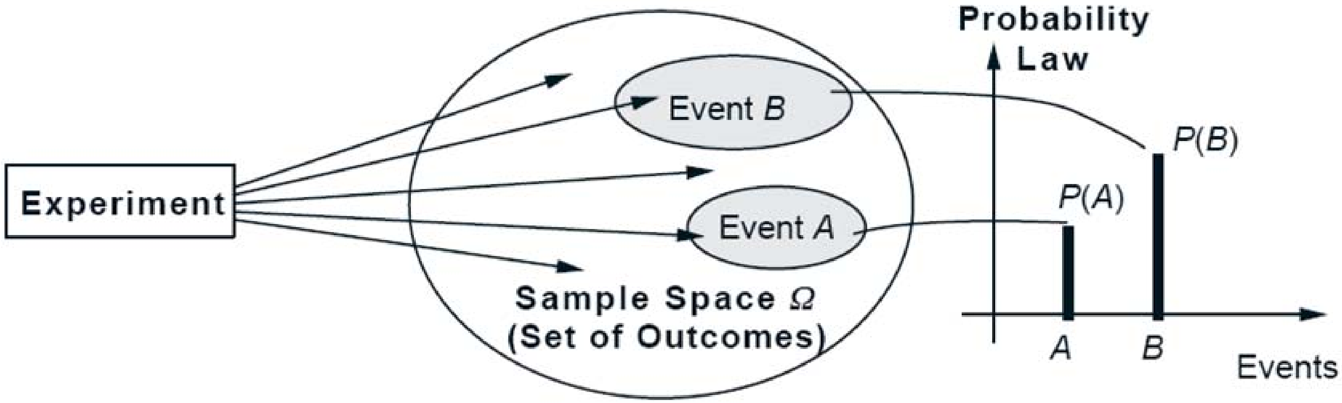
(a) **random variable** (r.v.): one random quantity

(b) **random sequence**: sequence of random variables

(c) **random process**: a (continuous-time) function whose value (at any time instant) is a r.v.

Probabilistic Model

- Random experiment: say, toss a coin
- **Outcome**: The output of conducting a random experiment; one at a time (exclusive).
- **Sample space**: The collection of all possible outcomes (of an experiment)
- **Probability law**: Probabilities assigned to the members of a sample space.



Probability Space Definition

- Relative Frequency -- experimental, intuitive
- Axiomatic Theory --- mathematical, rigorous, facilitate further derivation (*measure theory* in mathematics)
- *Example*: Tossing two fair coins

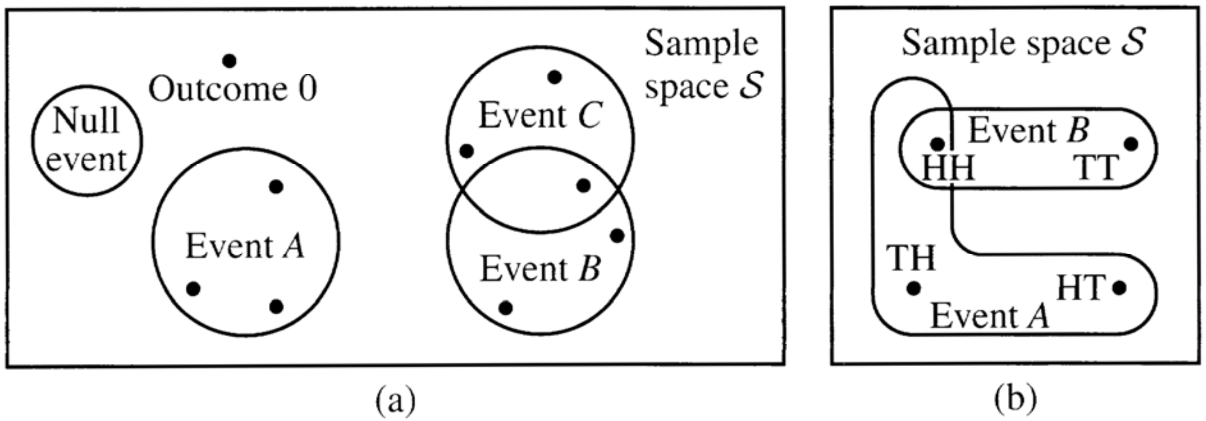


Figure 5.1
Sample spaces.(a) Pictorial representation of an arbitrary sample space. Points show outcomes; circles show events. (b) Sample space representation for the tossing of two coins.

Ex. Tossing Two Fair Coins

Relative Frequency	Axiomatic Theory
Model: $\begin{cases} Events: A \equiv HH, B \equiv HT \\ \text{Prob of } Events: P(A), P(B) \end{cases}$	Space: $\Omega = \{HH, HT, \cdots\}$ Field: $\Im = \{\{HH\}, \{HT\}, \cdots, \{HH \cup HT\}, \cdots, \varphi\}$
To find P(A) & P(B), we repeat experiment *N* times $\text{Prob}(A) \Box \lim_{N \to \infty} \frac{N_A}{N} = \frac{1}{4}$ $\text{Prob}(A \text{ or } B) \Box \lim_{N \to \infty} \frac{N_{A \cup B}}{N} = \frac{1}{2}$	Prob: (measure, mapping) $P : \Im \to [0,1]$ $P(\{HH\}) = 1/4$ $P(\{HH \cup HT\}) = 1/2$

Axioms of Probability

$\Omega =$ Sample space (a collection of elementary events)

$\Im =$ A collection of subsets of Ω and forms a filed (a σ-field)

Remarks:

(1) $(\Omega, \Im)$ is a field, if

(a) $\phi \in \Im$ and $\Omega \in \Im$.

(b) For any $\mathrm{A}, \mathrm{B} \in \Im$, $\mathrm{A} \cap \mathrm{B} \in \Im$.

(c) For any $\mathrm{A} \in \Im$, then $\mathrm{A}^c = \Omega - \mathrm{A} \in \Im$.

(2) σ-field: Consider countably infinite $\cup$ and $\cap$.

If $\mathrm{A}_i \in \Im$, then $\bigcup_{i=1}^{\infty} \mathrm{A}_i \in \Im$ and $\bigcap_{i=1}^{\infty} \mathrm{A}_i \in \Im$.

Probability Space

A probability space is a triplet $(\Omega, \Im, P)$.

(i) $(\Omega, \Im)$ is a σ-field.

(ii) P is a probability measure.

Remarks:

$P: \Im \to [0,1]$ is a probability measure on $\Im$ if

(a) For any $A \in \Im$, $P(A) \geq 0$.

(b) $P(\phi) = 0$ and $P(\Omega) = 1$.

(c) For any $A, B \in \Im$, if $A \cap B = \phi$, then $P(A \cup B) = P(A) + P(B)$.

(c') σ-additivity: For $A_i \in \Im$ and if $A_i \cap A_j = \phi$ for all $i \neq j$,

then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Example of Probability Space

- $\Omega = [0,1]$ line segment.
- $\Im = \{[0,1/2], (1/2,1], [0,1], \phi\}$
- *P*(A) = length of the line segment A.

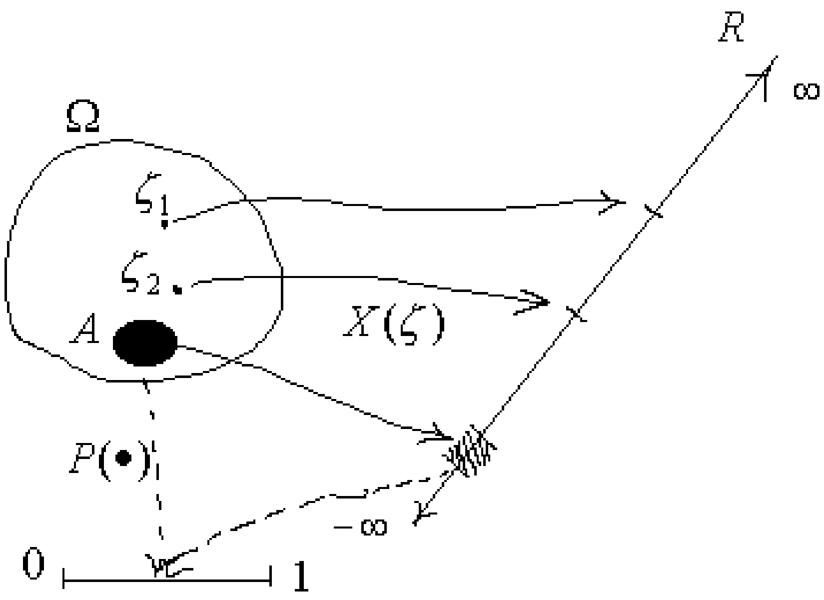
Random Variables

Definition: Given a probability space $(\Omega, \Im, P)$, a r.v. is a mapping $X : \Omega \to R$ such that

(i) the set $\{\zeta : X(\zeta) \in B\}$ must be a legal event A in $\Im$ for any Borel set B in R.

(ii) $P(X = -\infty) = P(X = \infty) = 0$.

Note: Borel set -- The smallest σ-field that contains all of the open sets in R. ($\approx$ all kinds of intervals on the real line.)



H Stock & J W Woods, *Prob. and Random Processes with Applications to Signal Proc.*, 3rd ed., Prentice-Hall, 2001

Remarks

(1) $(R,\{B\},P)$ is a derived probability space (from the original probability space $(\Omega,\Im,P)$).

If $X(\cdot)$ is properly defined (selected), $(R,\{B\},P)$ reflects all the probabilistic properties of $(\Omega,\Im,P)$. But $(R,\{B\},P)$ is often easier to handle (because of real line).

(2) A random variable $X(\cdot)$ is a function (mapping) not a simple value.

(3) Notations:

Capital letters $\leftrightarrow$ random variable: X, Y, Θ...

Lower-case letters $\leftrightarrow$ values of random variable: x, y, θ...

Probability Distribution Functions

Probability (Cumulative) Distribution Functions (PDF or cdf)

$F_X(x) \equiv P[X \le x]$, where $\{X \le x\} \equiv \{\zeta : X(\zeta) \le x\}$.

Properties:

(1) $F_X(-\infty) = 0, F_X(\infty) = 1$.

(2) Continuous from right: $\lim_{x \to x_{0^+}} F_X(x) = F_X(x_0)$.

(3) Nondecreasing: $F_X(x_1) \le F_X(x_2)$ if $x_1 \le x_2$.

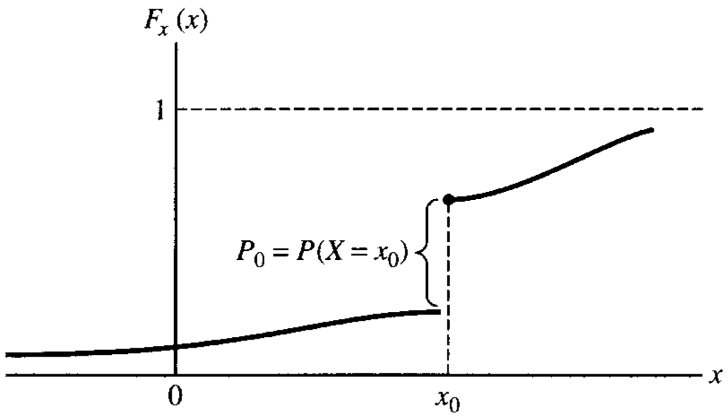


Figure 5.5
Illustration of the jump property of $F_X(x)$.

Probability Density Functions

Probability Density Functions (pdf)

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

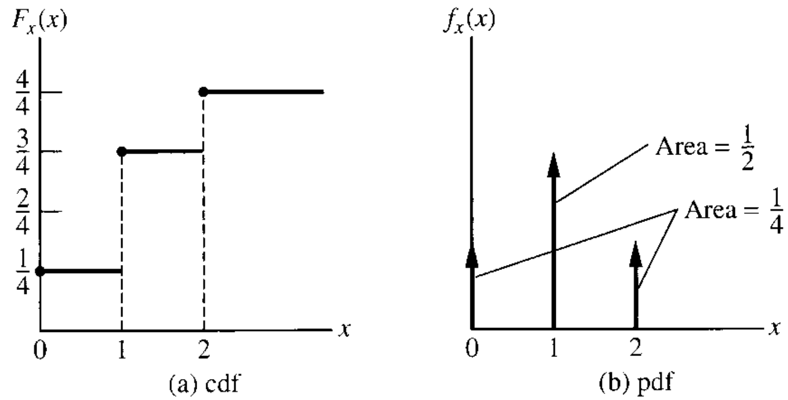


Figure 5.6
The cdf and pdf for a coin-tossing experiment.

Properties:

(1) $\int_{-\infty}^{\infty} f(\xi)d\xi = 1.$

(2) $F_X(x) = \int_{-\infty}^{x} f(\xi)d\xi = P[X \le x].$

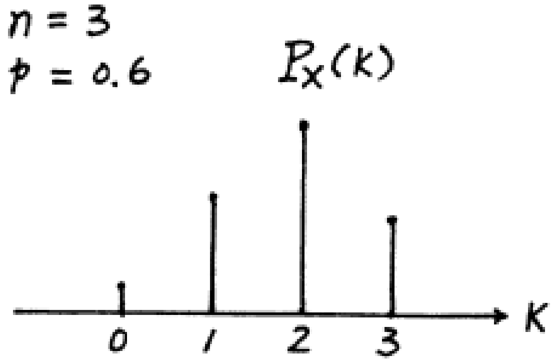
(3) $\int_{x_1}^{x_2} f(\xi)d\xi = F_X(x_2) - F_X(x_1) = P[x_1 < X \le x_2].$

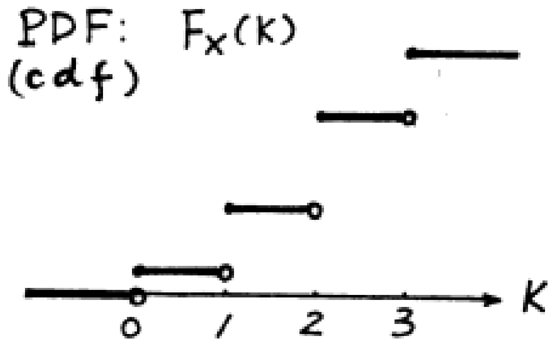
- Discrete Random Variable: *Probability mass function* (Prob of a single number can be non-zero.)

Useful pdfs

- Binomial distribution

$$P_X(k) = \binom{n}{k} p^k q^{n-k}, \quad \text{for } k = 0, 1, \ldots, n$$
$$= 0, \quad \text{otherwise.}$$





Remarks: the Laplace approximation to the binomial distribution

$$P_n(k) \cong \frac{1}{\sqrt{2\pi npq}} \exp[-\frac{(k-np)^2}{2npq}].$$

- Poisson Distribution

For $\alpha > 0$, $P_T(k) = \dfrac{(\alpha T)^k}{k!} e^{-\alpha T}$ for $k = 0, 1, 2, \ldots$

When n is large and p is small

$$P_n(k) \cong \frac{(\bar{K})^k}{k!} e^{-\bar{K}}, \text{ where } \bar{K} = E[K].$$

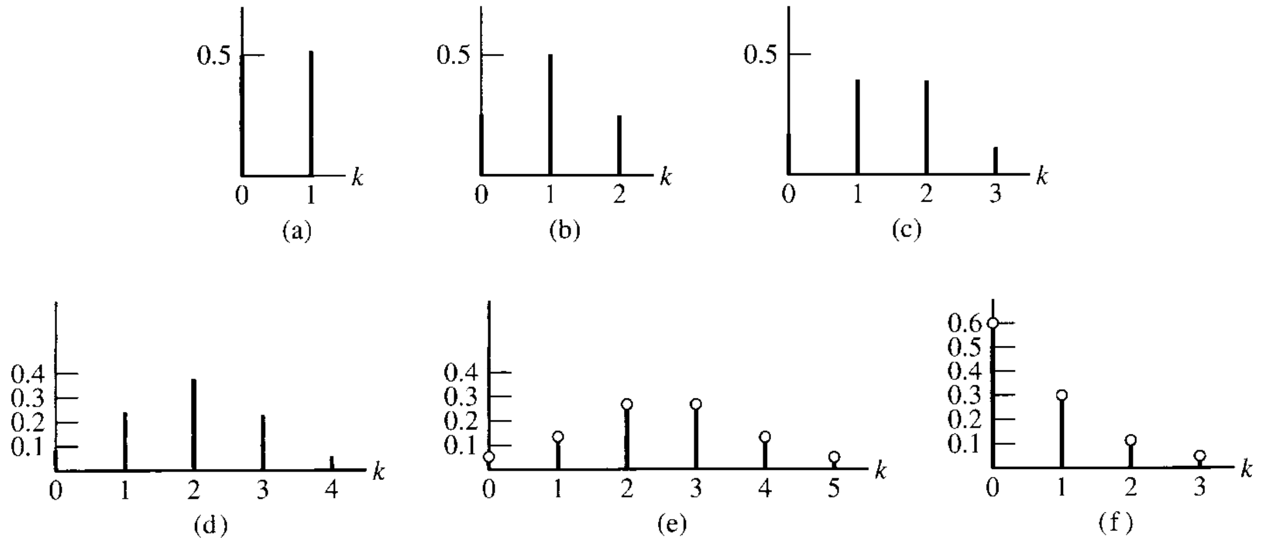
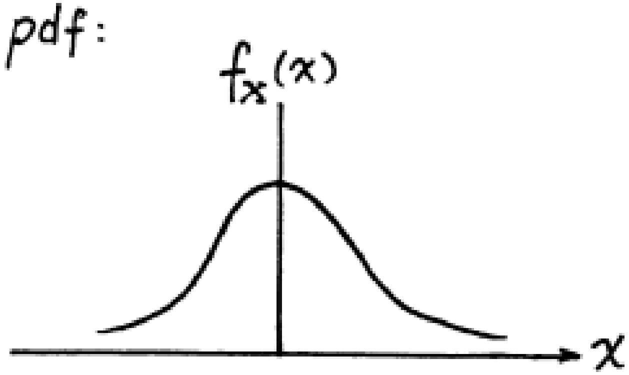
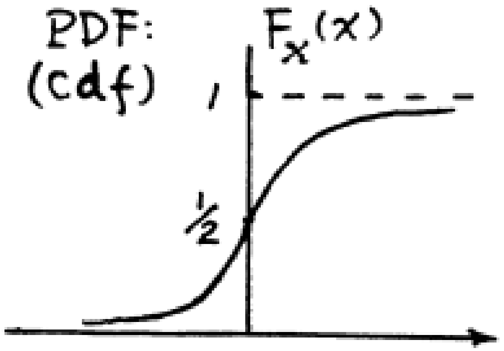


Figure 5.17
The binomial distribution with comparison to Laplace and Poisson approximations. (a) $n = 1, p = 0.5$. (b) $n = 2$, $p = 0.5$.(c) $n = 3, p = 0.5$. (d) $n = 4, p = 0.5$. (e) $n = 5, p = 0.5$. Circles are Laplace approximations. (f) $n = 5$, $p = \frac{1}{10}$. Circles are Poisson approximations.

- Gaussian (normal) distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}.$$





- Multi-dim Gaussian: $\mathbf{X} = (X_1, X_2, \cdots, X_n)$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(\det \mathbf{Q})^{1/2}} \times \exp\left[-\frac{1}{2}(\mathbf{x}-\beta)^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{x}-\beta)\right].$$

where $\mathbf{Q}$ is covariance matrix, β is mean vector.

Multiple Variables

- Joint cdf and pdf

$$F_{XY}(x, y) \triangleq P(X \leq x, Y \leq y).$$

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}.$$

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{XY}(x, y) dx dy.$$

Marginal distribution:

$$F_X(x) = F_{XY}(x, \infty) = F_{XY}(x, Y \leq \infty)$$

$$F_Y(y) = F_{XY}(\infty, y) = F_{XY}(X \leq \infty, y)$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy.$$

Marginal Prob

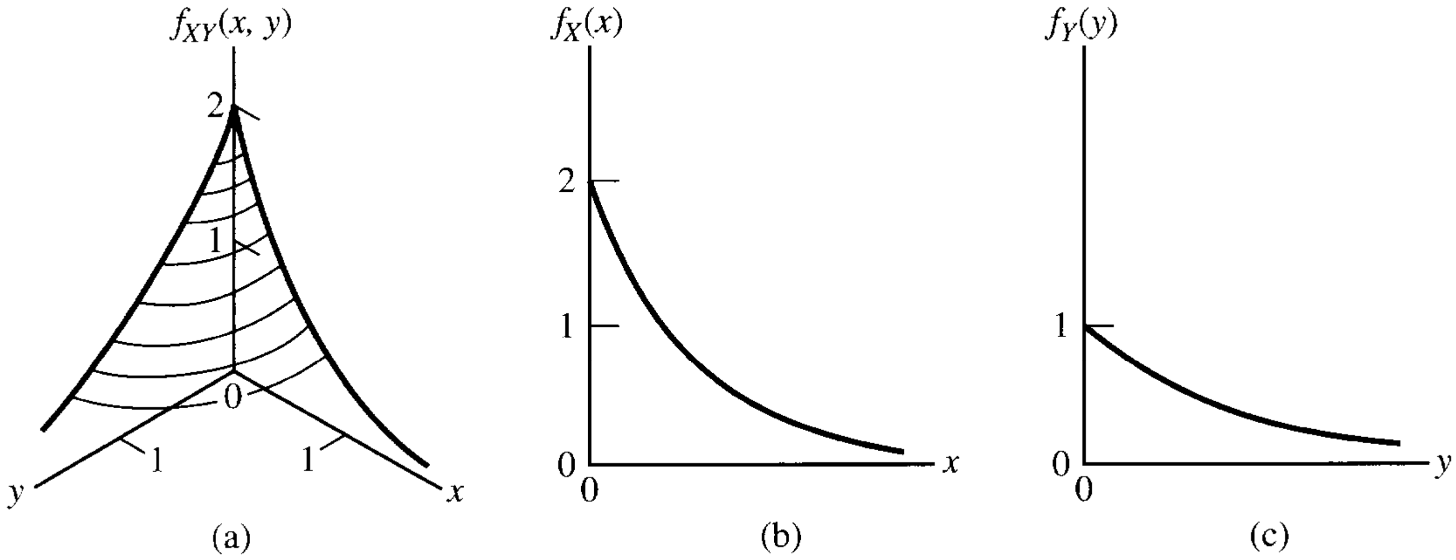
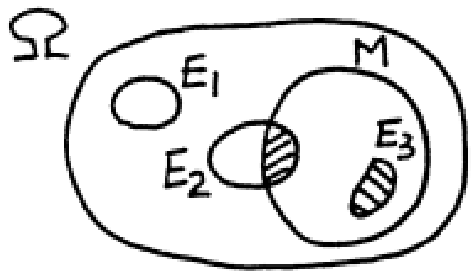


Figure 5.9
Joint and marginal pdfs for two random variables. (a) Joint pdf. (b) Marginal pdf for X. (c) Marginal pdf for Y.

- Conditional Probability: a derived probability measure

$$(\Omega, \mathcal{F}, P) \longrightarrow (\Omega, \mathcal{F}, P_M)$$

$$P_M[E] \equiv \frac{P[E \cap M]}{P[M]} \equiv P[E|M].$$



- Conditional cdf and pdf:

$$F(x|B) \equiv \frac{P[\{X \leq x\} \cap M]}{P[M]}.$$

$$f(x|B) \equiv \frac{dF(x|B)}{dx}.$$

- Conditional random variables:

$$F_{X|B}(x|B) = F_{X|Y}(x|Y \leq y) = \frac{F_{XY}(x,y)}{F_Y(y)}.$$

$$f_{X|Y}(x|y) = \frac{\partial F_{X|Y}(x|Y=y)}{\partial x} = \frac{f_{XY}(x,y)}{f_Y(y)}.$$

Bayes' Theorem

$$f_{X|Y}(x \mid y) = \frac{f_{Y|X}(y \mid x) f_X(x)}{f_Y(y)}$$

- Independent r.v.

$$P(X \le x, Y \le y) = P(X \le x)P(Y \le y).$$

$$F_{XY}(x, y) = F_X(x)F_Y(y).$$

$$f_{XY}(x, y) = f_X(x)f_Y(y).$$

Example: 2-D Gaussian

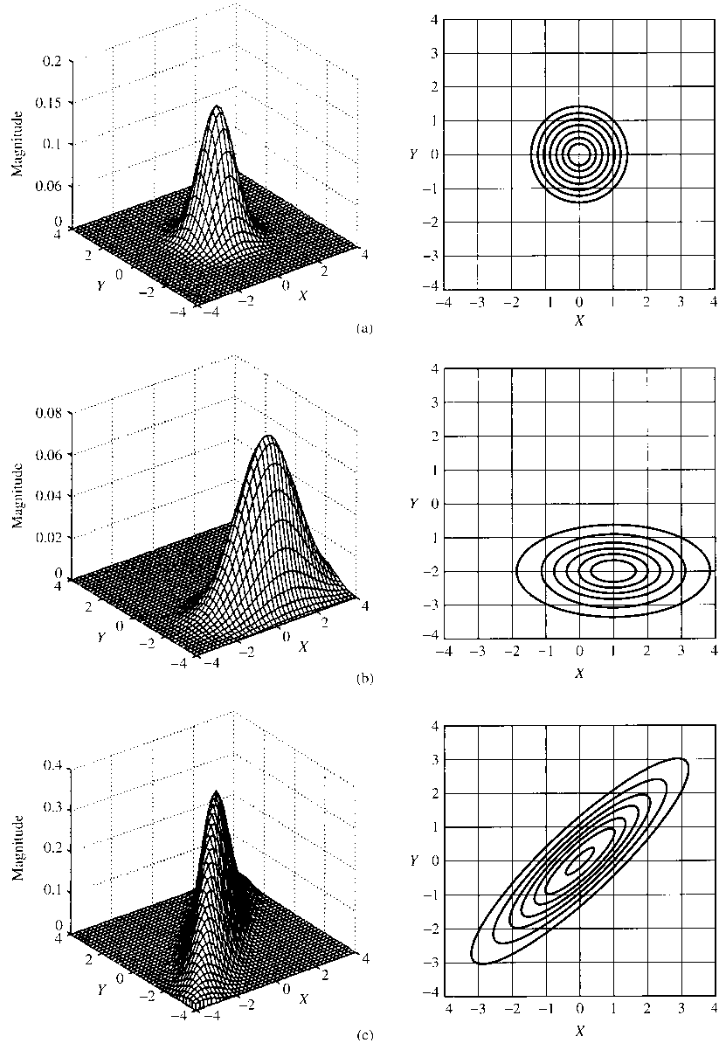


Figure 5.18
Bivariate Gaussian pdfs and corresponding contour plots. (a) $m_x = 0$, $m_y = 0$, $\sigma_x^2 = 1$, $\sigma_y^2 = 1$ and $\rho = 0$. (b) $m_x = 1$, $m_y = -2$, $\sigma_x^2 = 2$, $\sigma_y^2 = 1$, and $\rho = 0$. (c) $m_x = 0$, $m_y = 0$, $\sigma_x^2 = 1$, $\sigma_y^2 = 1$, and $\rho = 0.9$.

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$

$$\times \exp\left(\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right).$$

Bayes' Theorem

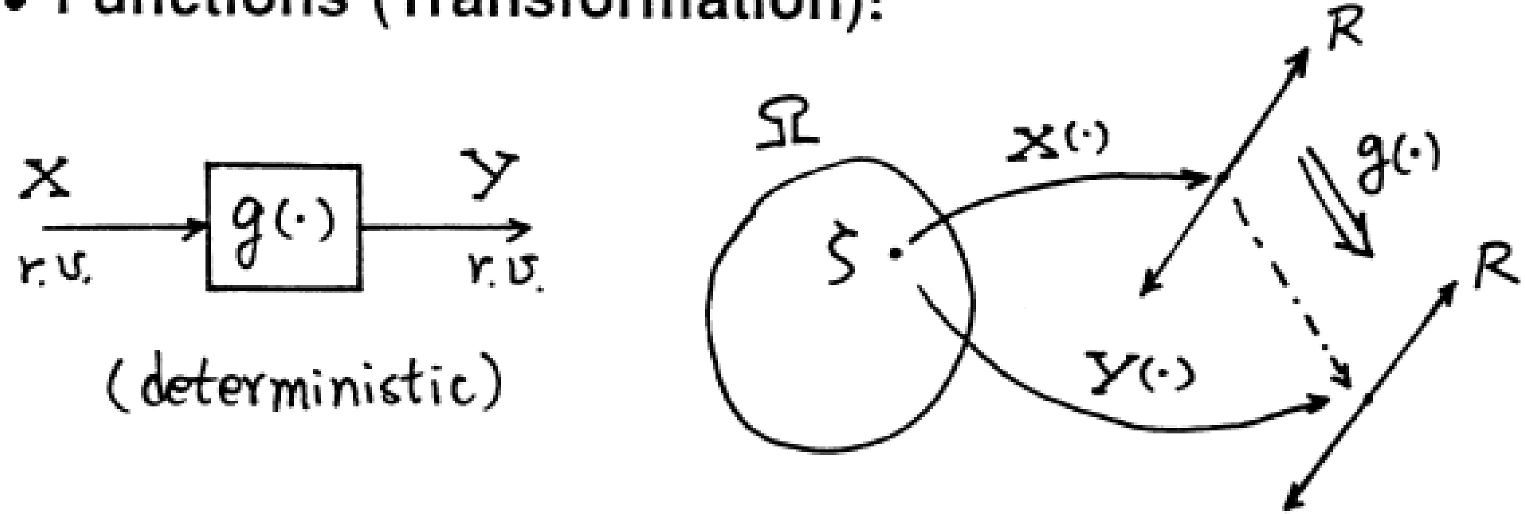
- Objective: Inference based on observations
- Given the observed “effect" or “result", infer the unobserved “cause". Assume we know the “prior" or “a priori" probabilities and the conditional probabilities in order to compute the “posterior" or “a posteriori" probabilities.

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}$$

Transformation of R.V.

- Create a new r.v. by transforming an old r.v.

- **Functions (Transformation):**



Basic: $\{\zeta : Y(\zeta) \leq y\} = \{\zeta : g(X(\zeta)) \leq y\} = \{\zeta : X(\zeta) \in C_Y\}$, where $C_Y = \{x : g(x) \leq y\}$.

Compute pdf

Assume $y = g(x)$ is a monotonically increasing function.

$$f_Y(y)|dy| = f_X(x)|dx|.$$

$$\Rightarrow f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|_{x=g^{-1}(y)}.$$

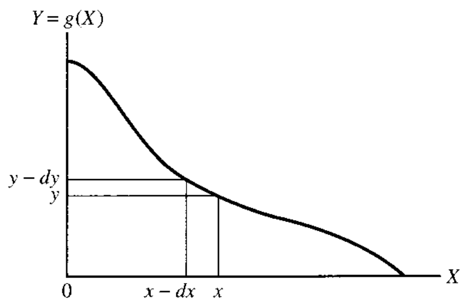


Figure 5.11
A typical monotonic transformation of a random variable.

In general, $f_Y(y) = \sum_{n=1}^{N} f_X(x_i)\left|\frac{dx_i}{dy}\right|_{x_i=g_i^{-1}(y)}.$

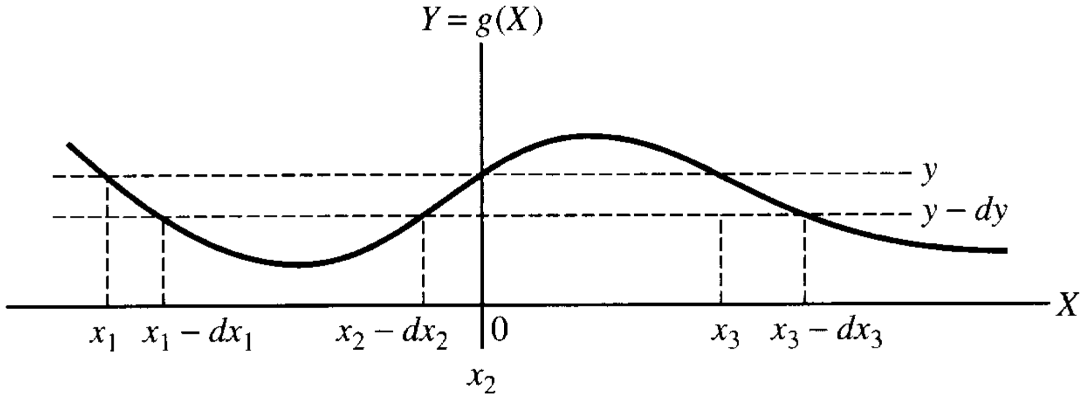


Figure 5.12
A nonmonotonic transformation of a random variable.

Multivariable Transformation

Let $\mathbf{X}$ be a random vector with pdf $f_{\mathbf{X}}(\mathbf{x})$ and $\mathbf{y} = g(\mathbf{x})$.

Assume the inverse exists, i.e., $\mathbf{x} = g^{-1}(\mathbf{y})$.

$$f_{\mathbf{Y}} d\mathbf{y} = f_{\mathbf{X}}(g^{-1}(\mathbf{y})) d\mathbf{x}. \Rightarrow f_{\mathbf{Y}} = f_{\mathbf{X}}(g^{-1}(\mathbf{y})) \frac{d\mathbf{x}}{d\mathbf{y}}.$$

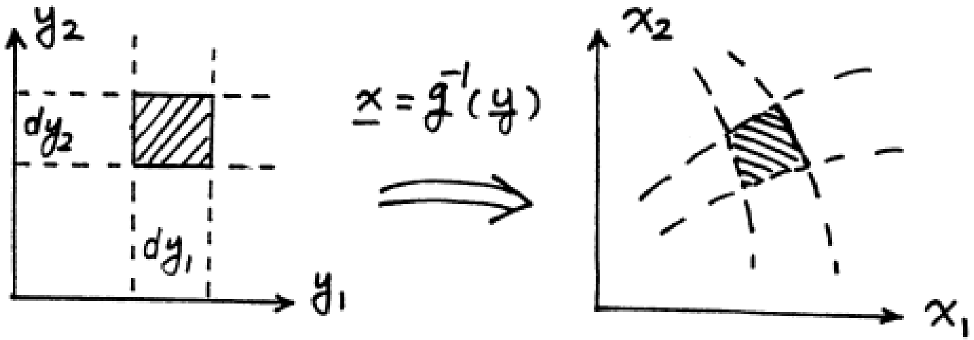
But $d\mathbf{x} = |\mathbf{J}| d\mathbf{y}$, where $\mathbf{J}$ is the Jacobian matrix.

$$\therefore f_{\mathbf{Y}} = f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |\mathbf{J}|.$$

$$f_{\mathbf{Y}} d\mathbf{y} = f_{\mathbf{X}} d\mathbf{x},$$

where $d\mathbf{y} = dy_1 dy_2 \ldots dy_n$.

$$\mathbf{J} = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix} = \text{Jacobian matrix.}$$



- *Example* 5.15: *X* and *Y* are independent and Gaussian, zero mean and variance=σ^2. Transform (X, Y) to (R, Θ).

$$f(x,y) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{x^2+y^2}{2\sigma^2}\right].$$

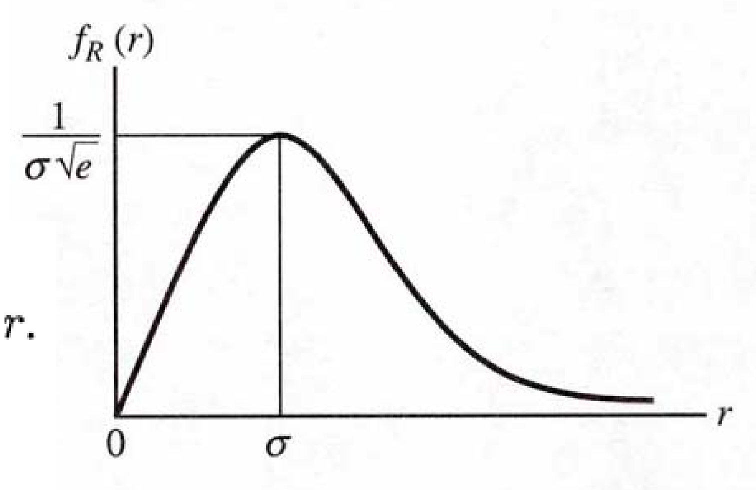
$$R = \sqrt{X^2+Y^2},\ \Theta = \tan^{-1}(Y/X)$$

$$X = R\cos\Theta,\ Y = R\sin\Theta.$$

$$\mathbf{J} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r cos^2\theta + r\sin^2\theta = r.$$

$$f_{R\Theta}(r,\theta) = \frac{r\exp[-r^2/(2\sigma^2)]}{2\pi\sigma^2},$$

$$f_R(r) = \int_0^{2\pi} f_{R\Theta}(r,\theta)d\theta = \frac{r}{\sigma^2}exp[-r^2/(2\sigma^2)].$$



Rayleigh pdf

Statistical Averages

- Mean (Weighted Average)

$$E[X] = \bar{X} \equiv \begin{cases} \sum_i x_i P_X(x_i), & \text{discrete r.v.;} \\ \int_{-\infty}^{\infty} x f(x) dx, & \text{continuous r.v.} \end{cases}$$

Remark: $E[\cdot]$ is a linear operator (of its arguement)

Additivity: $E[X+Y] = E[X] + E[Y]$ for two r.v.'s

Homogeneity: $E[cX] = cE[X]$ for any constant c

- The r-*th* moment, r = 0, 1, 2, …

$$\xi_r \equiv E[X^r] = \overline{X^r} = \begin{cases} \sum_i x_i^r P_X(x_i), & \text{discrete r.v.;} \\ \int_{-\infty}^{\infty} x^r f(x) dx, & \text{continuous r.v.} \end{cases}$$

- The r-*th* central momen $m_r \equiv E[(X - \bar{X})^r]$
  r = 0, 1, 2,
  Variance: $\sigma^2 \equiv m_2 = E[(X - \bar{X})^2] = E[X^2] - (\bar{X})^2.$

- The r-*th* joint moment, i, j = 0, 1, 2, …

$$\xi_{ij} \equiv E[X^i Y^j] = \begin{cases} \sum_i x_l^i y_m^j P_X(x_i, y_m), & \text{discrete r.v.;} \\ \int_{-\infty}^{\infty} x^i y^j f_{XY}(x,y)\,dx\,dy, & \text{continuous r.v} \end{cases}$$

- **Correlation**
  Note: Independent: $F_{XY}(x,y) = F_X(x)F_Y(y)$
  Uncorrelated: $E((X-E(X))(Y-E(Y))) = 0$
  Orthogonal: $E(XY) = 0$

- The r-*th* joint central moment

$$m_{ij} \equiv E[(X-\bar{X})^i(Y-\bar{Y})^j]$$

- Covariance

$$Cov[X,Y] \equiv m_{11} = E[(X-\bar{X})(Y-\bar{Y})] = E[XY]-\bar{X}\bar{Y}.$$

- Correlation coefficient

$$\rho \equiv \frac{m_{11}}{\sqrt{m_{20}m_{02}}}$$

$$Note\colon\ m_{02} = E[(Y-\bar{Y})^2],\ m_{20} = E[(X-\bar{X})^2];$$

$$-1 \le \rho \le 1$$

- Conditional expectation

$$E[X|B] = \int_{-\infty}^{\infty} x f_{X|B}(x|B)dx.$$

$$E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|Y=y)dx.$$

- Expectation of functions of r.v.: $Y = g(X)$

$$\bar{Y} = E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)dx.$$

- Moment generating functions

$$\theta(t) \equiv E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx.$$

$$\xi_r = \left.\frac{d^r(\theta(t))}{dt^r}\right|_{t=0}, \quad r = 0, 1, \ldots$$

- Characteristic functions (FT of pdf)

$$\Phi(\omega) = E[e^{j\omega X}] = \int_{-\infty}^{\infty} e^{j\omega x} f_X(x)dx.$$

- **Chebyshev Inequality:** X, an arbitrary r.v. with mean $\bar{X}$ and finite variance σ^2. Then, for any $\delta > 0$,

$$P[|X - \bar{X}| \geq \delta] \leq \frac{\sigma^2}{\delta^2}.$$

- **Law of Large Numbers:** Let $X_1, ..., X_n$ be i.i.d. (independent and identically distributed) r.v.'s with mean μ and variance σ^2 each. Let the **sample mean** be

$$\hat{\mu} \equiv \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Then, for any fixed $\delta > 0$,

$$\lim_{n \to \infty} P[|\hat{\mu} - \mu| \geq \delta] = 0.$$

- **Central Limit Theorem:** Let $X_1, ..., X_n$ be independent r.v.'s with zero mean and variance $\sigma_1^2, ..., \sigma_n^2$. Let $s_n^2 \equiv \sigma_1^2 + ... + \sigma_n^2$. If for any fixed $\epsilon > 0$, there exists a sufficient large n such that

$$\sigma_k^2 < \epsilon s_n, \text{for all } k = 1, ..., n,$$

then the normalized r.v.

$$Z_n \equiv (X_1 + ... + X_n)/s_n$$

converges to the standard normal (Gaussian) PDF.

Error Function and Q Function

Normalize a normal distribution $n(m_x, \sigma_x)$:

$$n(m_x, \sigma_x) \rightarrow n(0,1), \text{ i.e., } \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}} \rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$

Q function:

$$Q(u) \equiv \int_u^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \approx \frac{e^{-u^2/2}}{u\sqrt{2\pi}}, \quad \text{for } u >> 1.$$

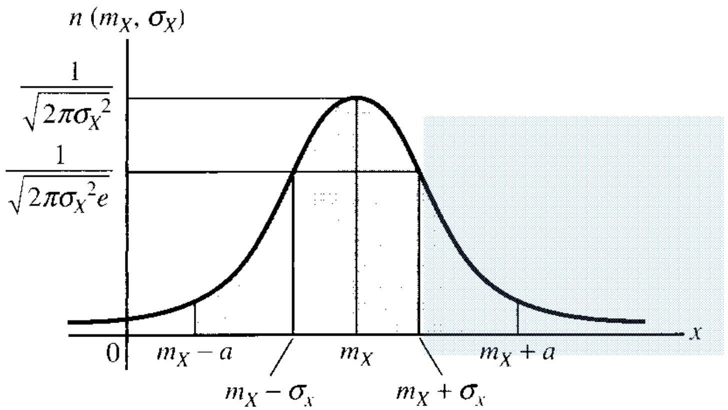


Figure 5.19
The Gaussian pdf with mean m_x and variance σ_x^2.

Q function is often used in the error prob analysis for digital communications. -->

Error function:

$$P[(m_x - a) \le X \le (m_x + a)] = 2[\frac{1}{2} - \int_{a/\sigma_x}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy] = 1 - 2Q(a/\sigma_x).$$

Define $erf(u) \equiv \frac{2}{\sqrt{\pi}} \int_0^u e^{-y^2} dy = 1 - 2Q(\sqrt{2}u).$

$$erf(u) = \frac{1}{\sqrt{\pi}} \int_{-u}^{u} e^{-y^2} dy \overset{y \equiv \frac{s}{\sqrt{2}}}{=} \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{2}u}^{\sqrt{2}u} e^{-s^2/2} ds.$$

Let $erfc(u) \equiv 1 - erf(u).$

Thus, $Q(u) = \frac{1}{2} erfc(\frac{u}{\sqrt{2}}).$

Table 5.4 Probability Distributions of Some Random Variables with Means and Variances

Probability density or mass function	Mean	Variance
Uniform: $f_X(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$
Gaussian: $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-m)^2}{2\sigma^2}\right)$	m	σ^2
Rayleigh: $f_R(r) = \frac{r}{\sigma^2} \exp\left(\frac{-r^2}{2\sigma^2}\right), \quad r \ge 0$	$\sqrt{\frac{\pi}{2}}\sigma$	$\frac{1}{2}(4-\pi)\sigma^2$
Laplacian: $f_X(x) = \frac{\alpha}{2} \exp(-\alpha\|x\|), \quad \alpha > 0$	0	$2/\alpha^2$
One-sided exponential: $f_X(x) = \alpha \exp(-\alpha x) u(x)$	$1/\alpha$	$1/\alpha^2$
Hyperbolic: $f_X(x) = \frac{(m-1)h^{m-1}}{2(\|x\|+h)^m}, \quad m > 3, h > 0$	0	$\frac{2h^2}{(m-3)(m-2)}$
Nakagami-m: $f_X(x) = \frac{2m^m}{\Gamma(m)} x^{2m-1} \exp(-mx^2), \quad x \ge 0$	$\frac{1 \times 3 \times \ldots \times (2m-1)}{2^m \Gamma(m)}$	$\frac{\Gamma(m+1)}{\Gamma(m)\sqrt{m}}$
Central chi-square (n = degrees of freedom):* $f_X(x) = \frac{x^{n/2-1}}{\sigma^n 2^{n/2} \Gamma(n/2)} \exp\left(\frac{-x}{2\sigma^2}\right)$	$n\sigma^2$	$2n\sigma^4$
Lognormal:† $f_X(x) = \frac{1}{x\sqrt{2\pi\sigma_y^2}} \exp\left(\frac{-(\ln x - m_y)^2}{2\sigma_y^2}\right)$	$\exp\left(m_y + 2\sigma_y^2\right)$	$\exp(2m_y + \sigma_y^2) \times [\exp \sigma_y^2 - 1]$
Binomial: $P_n(k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \ldots, n; \quad p + q = 1$	np	npq
Poisson: $P(k) = \frac{\lambda^k}{k!} \exp(-\lambda), \quad k = 0, 1, 2, \ldots$	λ	λ
Geometric: $P(k) = pq^{k-1}, \quad k = 1, 2, \ldots$	$1/p$	q/p^2

*$\Gamma(m)$ is the gamma function and equals $(m-1)!$ for m an integer.

†The lognormal random variable results from the transformation $Y = \ln X$, where Y is a Gaussian random variable with mean m_y and variance σ_y^2.