Ruda Zhang

Notes on Probability Theory for Engineers

December 29, 2013

Springer
Berlin Heidelberg New York
Hong Kong London
Milan Paris Tokyo

Contents

1

Preliminaries

Types of Probability + Analysis Methods:

1. Intuitive probability
   - sense from human experience.
   - Human are bad at making random lists.
2. Classical probability
   - simple, easy to understand.
   - cannot handle infinite number of outcomes, and assume all outcomes are equally likely.
3. Relative frequency
   Assumption (Statistical regularity): relative frequency approaches probability as n goes to infinity.

$$\lim_{n\to\infty} f_k(n) = p_k$$

   - agrees with experiment.
   - do not know how large should n (number of trial) be.
4. Abstract space (axiomatic) approach
   This is the approach we will use.

Definition 1. *A* ***probability space*** *is written as* $(\mathcal{U}, \mathcal{F}, \mathcal{P})$*, with:*

1. ***sample space*** $\mathcal{U}$*: the collection of all possible outcomes of a random experiment.* [1]
2. σ***-algebra*** $\mathcal{F}$*: the class of events of interest in a random experiment that forms a* σ*-algebra.* [2]
3. ***probability measure*** $\mathcal{P}$*: probability assignment of sets in* $\mathcal{F}$ *that satisfies the Axioms of Probability.*[3] [4]

[1] Sometimes we denote the sample space as Ω.
[2] $\mathcal{F}$ need not include all subsets of $\mathcal{U}$.
[3] In this sense, sets in $\mathcal{F}$ are said to be **measurable**. Any subset of $\mathcal{U}$ that is not in $\mathcal{F}$ is not measurable.
[4] Keep in mind that probability measure is a different concept from cdf/pdf.

Definition 2. ***σ-algebra*** *(also* ***σ-field****)* $\mathcal{F}$ *is a collection of subsets of a universe* $\mathcal{U}$*, which satisfies the following axioms:* [5]

1. $\mathcal{F}$ *is non empty:* $\exists A \subseteq \mathcal{U}, s.t. A \in \mathcal{F}$
2. $\mathcal{F}$ *is closed under complementation: if* $A \in \mathcal{F}$*, then* $A^c \in \mathcal{F}$
3. $\mathcal{F}$ *is closed under countable unions: if* $A_1, A_2, \cdots \in \mathcal{F}$*, then* $\bigcup_i A_i \in \mathcal{F}$

Definition 3. *Events* A_i *and* A_j *are* ***mutually exclusive****, iff* $A_i \cap A_j = \emptyset$*.*

Axioms of Probability (Kolmogorov Axiom):

1. $P(A) \geq 0, \forall A \in \mathcal{F}$
2. $P(\mathcal{U}) = 1$
3. Countable additivity: If $A_1, A_2, \cdots \in \mathcal{F}$, and they are all mutually exclusive, then

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$$

Definition 4. *An event from a discrete sample space that consists of a single outcome is called an* ***elementary event****.*

Definition 5. *The* ***Borel σ-algebra*** *is the σ-algebra generated by open sets in sample space* $\mathcal{U}$*. If the sample space is* $\mathbb{R}$*, the Borel σ-algebra is generated simply by open intervals of* $\mathbb{R}$*, and is denoted as* $\mathcal{B}^1$*.* [6]

Notation	Probability Space	Set theory
$\mathcal{U}$	Sample space	Universe
$u \in \mathcal{U}$	Outcome	Element
$A \in \mathcal{F}$	Event	Subset
$A_i \cap A_j = \emptyset$	mutually exclusive	disjoint

Table 1.1. Comparison of terms

1.1 Review of set theory

Definition 6. *Sets* A_i *and* A_j *are* ***disjoint****, iff* $A_i \cap A_j = \emptyset$*.*

Definition 7. *The* ***certain event****,* $\mathcal{U}$*, is the event consisting of all outcomes.*

Definition 8. *The* ***impossible (null) event****,* $\emptyset$*, is the event containing no outcomes.*

[5] There's a concept called smallest σ-algebra generated by a collection of subsets.
[6] The Borel σ-algebra contains all the subsets of $\mathcal{U}$ that engineers can use.

Definition 9. *A^c is the* ***complement*** *of set A.*

Definition 10. *The collection (also class) of all subsets of set S is called the* ***power set of*** *S, denoted by $\mathcal{P}(S)$.*[7]

1.2 Combinatorics

Assumptions of classical probability

1. $\mathcal{U}$ is finite
2. $\mathcal{F}$ = all subsets of $\mathcal{U}$
3. $\mathcal{P}$: all outcomes are equal in probability

Notation 1

$$(n)_k \equiv \frac{n!}{(n-k)!}$$

Definition 11. ***Permutation*** *is a reordering of a set of objects, the number of permutations of n objects is n!.*

Definition 12. ***Binomial coefficients***[8] *are (n choose k)*

$$\binom{n}{k} \equiv \frac{n!}{k!(n-k)!}$$

[7] From the axiom of power set, every set has a power set.

[8] A mathematical equation may have an interpretation apart from mathematical manipulation. Say

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$$

, apart from directly proving it mathematically, we can think it as picking k elements from $n+1$ elements. We can either pick them without one particular element, or with this element, which exhausted all possibilities. So the identity holds.

2

Set Probability Theory

Motivation for introducing **conditional probability**: Knowledge that one event has occurred may affect the probability that another will occur.

Definition 13. *The **conditional probability** of event A given that B has occurred is* [1]

$$P(A|B) \equiv \frac{P(A \cap B)}{P(B)}$$

Definition 14. *Events A and B are **statistically independent**[2] iff*

$$P(A \cap B) = P(A)\, P(B)$$

Definition 15. *A finite collection of events $\{A_1, A_2, \ldots, A_n\}$ is **mutually independent**[3] iff*

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) * \cdots * P(A_{i_k})$$

, $\forall \{A_{i_1}, \ldots, A_{i_k}\} \subseteq \{1, \ldots, n\}$

2.1 Total probability and Bayes' Rule

Definition 16. *A **partition** of the sample space $\mathcal{U}$ is mutually exclusive events whose union equals the sample space $\mathcal{U}$.*

Theorem 1. ***(Theorem of Total Probability)*** *Let $A_1, \ldots, A_n$ be a partition of $\mathcal{U}$, then*

$$P(B) = \sum_i P(B|A_i)\, P(A_i)$$

[1] Now we get a new probability space under conditional probability: $(\mathcal{U}, \mathcal{F}, \mathcal{P}(\cdot|B))$

[2] Two mutually exclusive events with nonzero probability can never be statistically independent.

[3] Pairwise independent events may not be mutually independent.

Theorem 2. ***(Bayes' Law)***

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Maximum a posteriori (MAP) estimation:
If $P(A_i|B) \geq P(A_j|B), \forall j \neq i$, then decide that A_i occurred.[4]

2.2 Binomial probability law

Definition 17. *If the events of separate experiments are mutually independent, we call these experiments as* ***independent experiments****.*

Theorem 3. *Let k be the number of successes in n independent Bernoulli trials, then the probability of k is given by the* ***binomial probability law***[5]*:*

$$P_n(k) = \binom{n}{k} p^k q^{n-k}$$

Poisson approximation:
If $n \gg 1$ and $np \sim 1$, denote $\lambda = np$, then

$$\begin{aligned} P_n(k) &= \binom{n}{k} p^k q^{n-k} \\ &= \frac{1}{k!}\frac{n!}{(n-k)!}\frac{\lambda^k}{n^k}(1-\frac{\lambda}{n})^n(1-\frac{\lambda}{n})^{-k} \\ &\approx \frac{\lambda^k}{k!}(1-\frac{\lambda}{n})^n \\ &\approx \frac{\lambda^k}{k!}e^{-\lambda} \end{aligned}$$

This gives the Poisson approximation[6]:

$$P_n(k) \approx \frac{\lambda^k}{k!}e^{-\lambda}$$

Gaussian approximation (de Moivre-Laplace theorem[7])
If $n \gg 1$, $npq \gg 1$ and k close to np, Gaussian approximation gives

$$P_n(k) \approx \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} N(x; np, npq)\,dx \approx N(k; np, npq)$$

---

4 This rule minimizes the probability of decision error, if we want to make a decision of which of $A_1, \ldots, A_n$ occurred based on the knowledge that B occurred.

5 Property: Define $k_{max} \equiv \arg\max_k P_n(k)$, then $k_{max} = \lfloor (n+1)p \rfloor$.

6 Poisson approximation is not reliable when $(1 - \sum_{k=0}^{n} \frac{\lambda^k}{k!} e^{-\lambda})$ is a probability of interest.

7 It's a special case of the central limit theorem.

Definition 18. ***Q-function*** *is defined as*

$$Q(z) \equiv \int_z^{\infty} N(x;0;1)\,dx$$

With Q-function, Gaussian approximation can be written as

$$P_n(k) \approx Q\left(\frac{k - np - \frac{1}{2}}{\sqrt{npq}}\right) - Q\left(\frac{k - np + \frac{1}{2}}{\sqrt{npq}}\right)$$

3

Random Variables

Motivation: We engineers deal with numbers and measurements that appears to be random.

Definition 19. ***Random variable (r.v.)***, *$X(u)$, is a function that maps an outcome of a random experiment to the real line ($X : \mathcal{U} \to \mathbb{R}$), with $P(E_z)$ defined, where event $E_z \equiv \{u \in \mathcal{U} | X(u) \le z\} \in \mathcal{F}, \forall z \in \mathbb{R}$.*[1] [2]

Definition 20. *For a fixed $u_0 \in \mathcal{U}$, $X(u_0) = x_0$ is called a* ***realization*** *of the r.v. X.*

Definition 21. *r.v.'s X and Y are* ***identically distributed****, if for every event $A \in \mathcal{B}^1$, $P(x \in A) = P(y \in A)$.*

Notation 2 $P_R\{X(u) \in B\} \equiv P(\{u \in \mathcal{U} | X(u) \in B\})$

3.1 Complete statistical description

Definition 22. ***Cumulative distribution function (cdf)*** *is defined as*

$$F_X(z) \equiv P(E_z), z \in \mathbb{R}$$

, where E_z is as the same as in Definition 19.

---

[1] It can be shown that $\forall B \in \mathcal{B}, X^{-1}(B) \in \mathcal{F}$. In this sense, random variable is a measurable function on the Borel σ-algebra.

[2] Random variable essentially makes the following transition:

$$(\Omega, \mathcal{F}, \mathcal{P}) \to (\mathbb{R}, \mathcal{B}, F_X(z))$$

Definition 23. ***Bernoulli r.v. (indicator r.v.)*** *is defined as*

$$X(u) = I_A(u) \equiv \begin{cases} 1, & u \in A \\ 0, & u \notin A \end{cases}$$

with cumulative distribution function

$$F_X(z) = (1-p)u(z) + pu(z-1)$$

, where $u(z)$ is the Heaviside step function.

Definition 24. ***Binomial r.v.*** *is the sum of n independent Bernoulli r.v.'s with $p, (1-p)$ each.*

Definition 25. ***Exponential r.v.*** *is defined as*[3]

$$F_X(z) = (1 - e^{-\lambda z})u(z) \quad (\lambda > 0)$$

Properties of cdf:

1. $0 \leq F_X(z) \leq 1, \forall z \in \mathbb{R}$
2. $\lim_{z\to\infty} F_X(z) = 1; \lim_{z\to-\infty} F_X(z) = 0$
3. $F_X(z)$ is non-decreasing
4. $1 - F_X(z) = P(E_z^c)$
5. $F_X(z)$ is continuous from the right[4]

$$F_X(z) = F_X(z+)$$

6. $P_R\{a < X \leq b\} = F_X(b) - F_X(a)$
7. $P_R\{X = b\} = F_X(b) - F_X(b-)$

Three types of r.v.'s:

1. **Continuous r.v.**: $F_X(z)$ is continuous, $\forall z \in \mathbb{R}$
2. **Discrete r.v.**: $X(u)$ takes a finite or countably many number of values $\{z_1, z_2, \ldots\}$
3. **Random variable of mixed type**: $X(u)$ takes some values with non-zero probability but also has probability distributed over the continuum.

Theorem 4. *r.v.'s X and Y are identically distributed $\iff$ $F_X(x) = F_Y(x), \forall x \in \mathbb{R}$.*

[3] Random variables are essentially (completely statistically) defined by their cdf.
[4] This can be easily seen since

$$F_X(z) = P_R\{X(u) \leq z\}$$

Definition 26. *Discrete r.v. has* ***probability mass function (pmf)****:*[5]

$$P_X(k) = P_R\{X = z_k\}$$

Definition 27. *Continuous r.v. has* ***probability density function (pdf)****:*[6]

$$f_X(z) = \frac{\mathrm{d}}{\mathrm{d}z} F_X(z)$$

Definition 28. ***Stieljes Integral*** *is*

$$\begin{aligned} P_R\{X(u) \in B\} &\equiv \int_{-\infty}^{\infty} I_B(x)\ \mathrm{d}F_X(x) \\ &= \int_B \mathrm{d}F_X(x) \end{aligned}$$

where

$$I_B(x) \equiv \begin{cases} 1, & x \in B \\ 0, & x \notin B \end{cases}$$

Notation 3 *If r.v.* X *has a distribution given by* $F_X(x)$*, we denote (****distributed as****)* $X \sim F_X(x)$*. Similarly, we can write* $X \sim f_X(x), X \sim Y$*.*

3.2 Complete statistical description conditioned on an event

Definition 29. ***Conditional cdf*** *is*

$$F_X(z|A) \equiv \frac{P(A \cap E_z)}{P(A)}$$

Definition 30. ***Conditional pdf*** *is*

$$f_X(z|A) \equiv \frac{\mathrm{d}}{\mathrm{d}z} F_X(z|A)$$

---

[5] Bernoulli r.v. has pmf:

$$P_X(k) = \begin{cases} 1-p, & k = 0 \\ p, & k = 1 \end{cases}$$

Binomial r.v. has pmf:

$$P_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \ldots, n$$

[6] Typically, we use delta functions for pdf of mixed type r.v.'s

Definition 31. ***Conditional pmf*** *is*

$$P_X(k|A) \equiv \frac{P_R\{X = k, A\}}{P(A)}$$

Notation 4

$$\begin{aligned} P(B|X = z) &\equiv \lim_{\Delta z \to 0+} \frac{P_R\{z < X \le z + \Delta z, B\}}{P_R\{z < X \le z + \Delta z\}} \\ &= \lim_{\Delta z \to 0+} \frac{\frac{[F_X(z+\Delta z|B) - F_X(z|B)]P(B)}{\Delta z}}{\frac{F_X(z+\Delta z) - F_X(z)}{\Delta z}} \\ &= \frac{f_X(z|B)P(B)}{f_X(z)} \end{aligned}$$

Theorem 5. *Two* ***"mixed forms" of Bayes' Law****:*

$$f_X(x|B) = \frac{P(B|X = x)\ f_X(x)}{P(B)} \tag{3.1}$$

with

$$P(B) = \int_{-\infty}^{\infty} P(B|X = x)\ f_X(x)\ \mathrm{d}x$$

And given $\{A_1, A_2, \ldots, A_n\}$ *is a partition of sample space,*

$$P(A_i|X = x) = \frac{f_X(x|A_i)P(A_i)}{f_X(x)} \tag{3.2}$$

with

$$f_X(x) = \sum_{i=1}^{n} f_X(x|A_i)P(A_i)$$

3.3 One r.v. as a deterministic function of another

Definition 32. ***Function of a r.v.*** *is a r.v., as illustrated in Figure 3.1:*

$$Y(u) \equiv g(X(u))$$

Cdf of Y *is expressed as*

$$\begin{aligned} F_Y(y) &\equiv P_R\{Y \le y\} \\ &= P_R\{g(X(u)) \le y\} \\ &= P_R\{X(u) \in G_y\} \end{aligned}$$

where $G_y \equiv \{x \in \mathbb{R} | g(x) \le y\}$ *is the inverse image of* $(-\infty, y]$ *under* $g(\cdot)$*.*

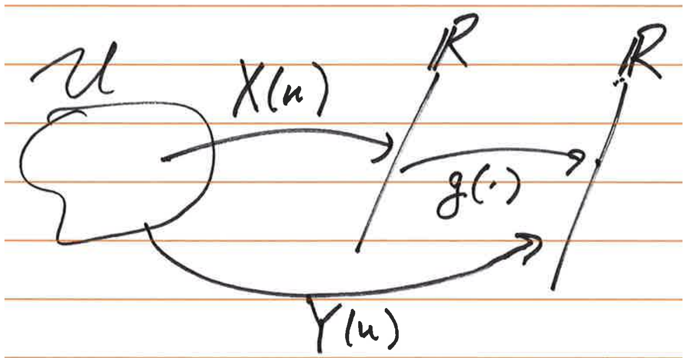

Fig. 3.1. Function of a random variable

Two cases of $f_Y(y)$:

1. $g(\cdot)$ is invertible on the range of $X(u)$.
   - $g(x) = ax + b \quad (a \neq 0)$,

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

   - $g(x) = e^x$

$$f_Y(y) = \frac{1}{y} f_X(\ln y) u(y)$$

2. $g(\cdot)$ is not uniquely invertible.
   - $g(x) = x^2$

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})]u(y)$$

Handling point masses for function of r.v.:
Point masses occur when there is:

1. point mass in $f_X(x)$;
2. flat region in $g(x)$;

We calculate point masses separately.
General formula for pdf of function of r.v.:
If $f_X(x)$ has no point mass, and $g(x)$ is differentiable and has no flat region, pdf of r.v. $Y(u) = g(X(u))$ is:

$$f_Y(y) = \sum_{\{i|g(x_i)=y\}} \frac{f_X(x_i)}{|g'(x_i)|}$$

3.4 Incomplete statistical descriptions

Definition 33. *The* ***expectation operator*** *(ensemble average) is defined as:*

$$\mathbb{E}\{X(u)\} \equiv \int_{-\infty}^{+\infty} x \, \mathrm{d}F_X(x)$$

The expectation operator for function of r.v. is:

$$\mathbb{E}\{g(X(u))\} \equiv \int_{-\infty}^{+\infty} g(x) \, \mathrm{d}F_X(x)$$

Properties of expectation operator:

1. Linearity:

$$\mathbb{E}\{aX(u) + bY(u)\} = a\mathbb{E}\{X(u)\} + b\mathbb{E}\{Y(u)\}$$

2. Jensen's Inequality: If $g(x)$ is convex,

$$\mathbb{E}\{g(X(u))\} \geq g(\mathbb{E}\{X(u)\})$$

Definition 34. *The* ***mean*** *of $X(u)$ is defined as:*

$$m_x \equiv \int_{-\infty}^{+\infty} x \, \mathrm{d}F_X(x)$$

It can be seen that: $m_x = \mathbb{E}\{X(u)\}$.

For continuous r.v.:

$$m_x = \int_{-\infty}^{+\infty} x f_X(x) \, \mathrm{d}x$$

For discrete r.v.:

$$m_x = \sum_k x_k P_X(k)$$

Definition 35. *The* ***variance*** *of $X(u)$ is defined as:*

$$\sigma_x^2 \equiv \int_{-\infty}^{+\infty} (x - m_x)^2 \, \mathrm{d}F_X(x)$$

, where σ_x is called the ***standard deviation of*** *$X(u)$.*[7]

It can be seen that

$$\sigma_x^2 = \mathbb{E}\{(X(u) - m_x)^2\} = \mathbb{E}\{X(u)^2\} - m_x^2$$

[7] The mean and/or variance of a r.v. may not exist (not finite), e.g.

$$f_X(x) = \frac{\alpha}{\pi} \frac{1}{x^2 + \alpha^2}$$

Definition 36. *m_x is called the 1^{st}* ***moment of*** *X; σ_x^2 is called 2^{nd}* ***central moment of*** *X; (m_x, σ_x^2) is called 2^{nd}* ***moment description of*** *X.*

Definition 37. *The n^{th}* ***(noncentral) moment of*** *X is defined as:*[8] [9]

$$\mathbb{E}\{X^n\} \equiv \int_{-\infty}^{+\infty} x^n f_X(x)\ \mathrm{d}x$$

The n^{th} ***central moment of*** *X is defined as:*[10]

$$\mathbb{E}\{(X - m_x)^n\} \equiv \int_{-\infty}^{+\infty} (x - m_x)^n f_X(x)\ \mathrm{d}x$$

3.5 Alternate complete statistical description

Definition 38. ***Characteristic function (cf) of*** *X is defined as:*

$$\Phi_X(\omega) \equiv \mathbb{E}\{e^{j\omega X}\} = \int_{-\infty}^{+\infty} e^{j\omega x} f_X(x)\ \mathrm{d}x \quad (j \equiv \sqrt{-1})$$

, which is the inverse Fourier transform of $f_X(x)$.[11]

Definition 39. ***Moment generating function (mgf)*** *of X is defined as:*

$$G_X(s) \equiv \mathbb{E}\{e^{sX}\} = \int_{-\infty}^{+\infty} e^{sx} f_X(x)\ \mathrm{d}x$$

, which is the two-sided Laplace transform of $f_X(-x)$, for s in some neighborhood of 0. If this expectation does not exist in any neighborhood of 0, we say that the moment generating function does not exist. [12]

---

[8] Higher order moments refine the incomplete description provided by m_x, σ_x^2.

[9] Tips:

$$\int_{0}^{+\infty} x^t e^{-x}\ \mathrm{d}x = t! \quad (t \in \mathbb{N})$$

[10] All moments of a Gaussian r.v. are determined by m_x and σ_x^2; odd central moments of a Gaussian r.v. are zero.

[11] The characteristic function always exist and the pdf can be recovered from $\Phi_X(\omega)$ with Fourier transform:

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-j\omega x} \Phi_X(\omega)\ \mathrm{d}\omega$$

[12] We can see that

$$\Phi_X(\omega) = G_X(j\omega)$$

Theorem 6. *If X has mgf $G_X(s)$, then for all n, the n-th moment of X exists and*

$$\mathbb{E}\{X^n\} = \frac{\mathrm{d}^n}{\mathrm{d}s^n} G_X(s)\bigg|_{s=0}$$

Theorem 7. *If r.v. X have bounded support and all the moments, then given n^{th} (noncentral) moments of X for $n = 1, 2, 3, \dots,$ we have a complete statistical description of X.*[13]

Theorem 8. *If r.v. X have all the moments and its mgf $G_X(s)$ exists, then given the mgf $G_X(s)$ for s in some neighborhood of 0, or all the moments, we have a complete statistical description of X.*[14]

Theorem 9. ***(Convergence of mgfs)*** *Suppose $\{X_i, i = 1, 2, \dots\}$ is a sequence of r.v.'s, each with mgf $G_{X_i}(s)$. Furthermore, suppose that*

$$\lim_{i\to\infty} G_{X_i}(s) = G_X(s)$$

, for all t in a neighborhood of 0, and $G_X(s)$ is an mgf. Then there is a unique cdf $F_X(x)$ whose moments are determined by $G_X(s)$ and, for all x where $F_X(x)$ is continuous, we have

$$\lim_{i\to\infty} F_{X_i}(x) = F_X(x)$$

Definition 40. *For discrete r.v. $X(u) \in \{0, \pm1, \pm2, \dots\}$,* ***probability generating function (pgf)*** *is defined as:*

$$\Gamma_X(z) \equiv \mathbb{E}\{z^{X(u)}\} = \sum_{k=-\infty}^{+\infty} P_X(k) z^k$$

We can see that

$$\Phi_X(\omega) = \Gamma_X(e^{j\omega})$$

Theorem 10. ***(moment theorem)***[15]

$$\frac{\mathrm{d}^k}{\mathrm{d}z^k} \Gamma_X(z)\bigg|_{z=1} = \mathbb{E}\{X \cdot (X-1) \cdots (X-k+1)\}$$

---

[13] See Billingsley 1995, Section 30 for proof.

[14] The mgfs fail to numerically distinguish the distributions, the characteristic functions do a fine job. (Waller et al. 1995, Luceño 1997)

[15] Proof unknown.

3.6 Tail probability bounds from moments

Markov bound [16] For non-negative r.v. $X(u)$ and $a > 0$,

$$P_R\{X \geq a\} \leq \frac{\mathbb{E}\{X(u)\}}{a}$$

Chebyshev's inequality (bound) [17]

$$P_R\{\left|X(u) - m_x\right| \geq \varepsilon\} \leq \frac{\sigma_x^2}{\varepsilon^2}$$

Non-central Chebychev

$$P_R\{\left|X(u)\right| \geq \varepsilon\} \leq \frac{\mathbb{E}\{X(u)^2\}}{\varepsilon^2}$$

Higher order tail bounds For $k = 1, 2, 3, \ldots$

$$P_R\{\left|X(u) - m_x\right| \geq \varepsilon\} \leq \frac{\mathbb{E}\{(X(u) - m_x)^{2k}\}}{\varepsilon^{2k}}$$

[16] Proof:

$$\begin{aligned}
\mathbb{E}\{X(u)\} &= \int_{-\infty}^{+\infty} x f_X(x)\ \mathrm{d}x \\
&= \int_{0}^{+\infty} x f_X(x)\ \mathrm{d}x \\
&\geq \int_{a}^{+\infty} x f_X(x)\ \mathrm{d}x \\
&\geq a \int_{a}^{+\infty} f_X(x)\ \mathrm{d}x \\
&= a P_R\{X \geq a\}
\end{aligned}$$

[17] Proof:

$$\begin{aligned}
P_R\{\left|X(u) - m_x\right| \geq \varepsilon\} &= P_R\{(X(u) - m_x)^2 \geq \varepsilon^2\} \\
&\leq \frac{\mathbb{E}\{(X(u) - m_x)^2\}}{\varepsilon^2} \\
&= \frac{\sigma_x^2}{\varepsilon^2}
\end{aligned}$$

Chernoff bound [18] [19]

$$P_R\{X \geq a\} \leq \min_{0<s<\infty} e^{-sa} G_X(s)$$

The "3σ rule of thumb":

With no less than 90% probability, a realization of X will be within 3σ of m_x.

[18] Proof:

$$\begin{aligned} P_R\{X(u) \geq a\} &= P_R\{sX(u) \geq sa\} \quad (s > 0) \\ &= P_R\{e^{sX(u)} \geq e^{sa}\} \\ &\leq \frac{\mathbb{E}\{e^{sX(u)}\}}{e^{sa}} \\ &= e^{-sa} G_{X(u)}(s) \end{aligned}$$

[19] This bound requires the complete statistical description, but sometimes it provides a very tight bound in simple form.

4

Pairs (and Finite Collections) of Random Variables

Definition 41. ***Pair of random variables****,* $\mathbf{X} = (X, Y)$*, is a function that maps an outcome of a random experiment to the real plane* ($\mathbf{X} : S \to \mathbb{R}^2$)*.*

4.1 Complete statistical descriptions

Definition 42. ***Joint cdf*** *of random variable* X *and* Y *is defined as*

$$F_{X,Y}(x, y) \equiv P_R\{X \le x, Y \le y\} \quad (x, y \in \mathbb{R})$$

Properties of joint cdf:

1. Nondecreasing: if $x_1 \le x_2, y_1 \le y_2$, then

$$F_{X,Y}(x_1, y_1) \le F_{X,Y}(x_2, y_2)$$

2.

$$\lim_{x \to -\infty} F_{X,Y}(x, y) = 0$$
$$\lim_{y \to -\infty} F_{X,Y}(x, y) = 0$$
$$\lim_{x,y \to +\infty} F_{X,Y}(x, y) = 1$$

3. Relation with **marginal cdf's** [1]

$$F_X(x) = \lim_{y \to +\infty} F_{X,Y}(x, y) \qquad F_Y(y) = \lim_{x \to +\infty} F_{X,Y}(x, y)$$

4. Continuous from right in each coordinate:

$$F_{X,Y}(x+, y) = F_{X,Y}(x, y) \qquad F_{X,Y}(x, y+) = F_{X,Y}(x, y)$$

[1] In this sense, $F_X(x)$ and $F_Y(y)$ are called marginal cdf's.

5. Probability in a rectangular area:

$$P_R\{x_1 < X \le x_2, y_1 < Y \le y_2\} = \\ F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1)$$

Definition 43. *Random variables X and Y are **jointly continuous**, if the probabilities of events involving (X, Y) can be expressed as an integral of a probability density function, i.e.*

$$P_R\{x \in B\} = \iint_B f_{X,Y}(x', y')\, \mathrm{d}x'\, \mathrm{d}y', \quad \forall B \in \mathcal{B}$$

Definition 44. *If random variables X and Y are jointly continuous random variables, then the **joint pdf** is*

$$f_{X,Y}(x, y) \equiv \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}$$

Properties of joint pdf:

1. Nonnegativity
$$f_{X,Y}(x, y) \ge 0$$
2. Relation with joint cdf
$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x', y')\, \mathrm{d}x'\, \mathrm{d}y'$$
3. Relation with **marginal pdf's**
$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)\, \mathrm{d}y$$
$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)\, \mathrm{d}x$$

Definition 45. *For a pair of discrete random variables $\mathbf{X} = (X, Y)$, the **joint pmf** of $\mathbf{X}$ is defined as*

$$P_{X,Y}(x_j, y_k) \equiv P_R\{X = x_j, Y = y_k\} \qquad (x_j, y_k) \in S_{X,Y}$$

4.2 Conditional distributions

Definition 46. ***(Conditional cdf)*** *The cdf of r.v. Y given the realization of r.v. $X = x$ is defined as*

$$F_{Y|X}(y|x) \equiv \lim_{\Delta x \to 0+} \frac{P_R\{x < X \leq x + \Delta x, Y \leq y\}}{P_R\{x < X \leq x + \Delta x\}}$$
$$= \lim_{\Delta x \to 0+} \frac{\int_{-\infty}^{y} f_{X,Y}(x, y')\ \mathrm{d}y' \Delta x}{f_X(x)\Delta x}$$
$$= \frac{\int_{-\infty}^{y} f_{X,Y}(x, y')\ \mathrm{d}y'}{f_X(x)}$$

, where $f_X(x) \neq 0$.

Definition 47. ***(Conditional pdf)*** *The pdf of r.v.* Y *given the realization of r.v.* $X = x$ *is defined as*[2]

$$f_{Y|X}(y|x) \equiv \tfrac{\partial}{\partial y} F_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

Theorem 11. ***(Total probability for continuous r.v.'s)***

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{Y|X}(y|x) f_X(x)\ \mathrm{d}x$$

Theorem 12. ***(Bayes law for continuous r.v.'s)***

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

4.3 Incomplete statistical descriptions

Definition 48. ***Joint moments*** *of two random variables* X, Y *has the general form of*

$$\mathbb{E}g(X,Y) = \iint_{-\infty}^{+\infty} g(x,y) f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{4.1}$$

For random vectors, it is

$$\mathbb{E}g(\mathbf{X}) = \int_{\mathbb{R}^n} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x} \tag{4.2}$$

Definition 49. ***Correlation*** *of a pair of random variables is their joint non-central* 2^{nd} *moment.*

$$COR[X,Y] \equiv \mathbb{E}XY \tag{4.3}$$

[2] From the equation, we can see that

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

Hence, $f_{Y|X}(y|x)$ and $f_X(x)$ is another complete statistical description for the pair of r.v.'s $\mathbf{X} = (X, Y)$.

Definition 50. ***Covariance*** *of a pair of random variables is their joint central 2^{nd} moment.* [3]

$$COV[X,Y] \equiv \mathbb{E}(X - m_x)(Y - m_y) \tag{4.4}$$

Definition 51. *Random variables X, Y are* ***orthogonal****, iff $COR[X,Y] = 0$; they are* ***uncorrelated****, iff $COV[X,Y] = 0$.*

Definition 52. ***Correlation coefficient*** *of a pair of random variables is* [4]

$$\rho_{XY} \equiv \frac{COV[X,Y]}{\sigma_X \sigma_Y} \tag{4.5}$$

Definition 53. *The* ***covariance matrix*** *of a random vector is*

$$K_{\mathbf{x}} \equiv \mathbb{E}[\mathbf{X}_0 \mathbf{X}_0^T] \tag{4.6}$$

Definition 54. *The* ***correlation matrix*** *of a random vector is*

$$R_{\mathbf{x}} \equiv \mathbb{E}[\mathbf{X}\mathbf{X}^T] \tag{4.7}$$

Definition 55. *The* ***second moment description*** *of a pair of random variables is the means m_x, m_y, variances σ_x^2, σ_y^2, and covariance $COV[X,Y]$.*

The ***second moment description*** *of a random vector is its mean vector $\mathbf{m}_{\mathbf{x}}$ and covariance matrix $K_{\mathbf{x}}$.*

4.4 Independent random variables

Definition 56. *X and Y are* ***statistically independent random variables****, iff any event A_1 defined in terms of X is independent of any event A_2 defined in terms of Y. i.e.,*

$$P_R\{X \in A_1, Y \in A_2\} = P_R\{X \in A_1\}\ P_R\{Y \in A_2\} \qquad \forall A_1, A_2 \subseteq \mathbb{R}$$

We denote it as $X \perp\!\!\!\perp Y$.

Theorem 13. ***(Alternate definition of independent r.v.'s)***

Random variables X and Y are independent iff their joint cdf is equal to the product of its marginal cdf's. i.e.,

$$X \perp\!\!\!\perp Y \iff F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y) \qquad \forall x, y \in \mathbb{R}$$

If X and Y are jointly continuous, then X and Y are independent iff their joint pdf is equal to the product of its marginal pdf's. i.e.,

$$X \perp\!\!\!\perp Y \iff f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \qquad \forall x, y \in \mathbb{R}$$

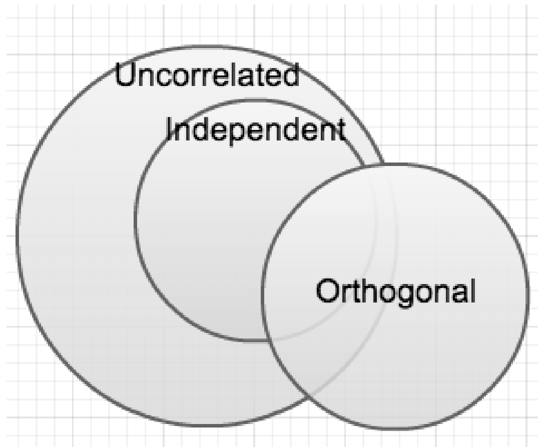

Fig. 4.1. Relations of uncorrelated, independent, and orthogonal random variables

Theorem 14. *If $X \perp\!\!\!\perp Y$, then they are uncorrelated. The converse doesn't hold.*

Theorem 15. *Let (X, Y) be a random vector.*

1. *If $X \perp\!\!\!\perp Y$, then $\forall g(\cdot), h(\cdot)$, $g(X) \perp\!\!\!\perp h(Y)$.*
2. *If $\forall g(\cdot), h(\cdot)$, $g(X)$ and $h(Y)$ are uncorrelated, then $X \perp\!\!\!\perp Y$*

Theorem 16. *Let X and Y be independent r.v.'s with characteristic functions $\Phi_X(\omega)$ and $\Phi_Y(\omega)$. Then the characteristic function of r.v. $Z = X + Y$ is*

$$\Phi_Z(\omega) = \Phi_X(\omega)\Phi_Y(\omega)$$

4.5 Alternate complete statistical description

Definition 57. ***Joint characteristic function*** *of a pair of random variables is* [5]

$$\Phi_{X,Y}(\omega_x, \omega_y) \equiv \mathbb{E}e^{j(\omega_x X + \omega_y Y)} \tag{4.8}$$

Theorem 17.

$$X \perp\!\!\!\perp Y \iff \Phi_{X,Y}(\omega_x, \omega_y) = \Phi_X(\omega_x) \cdot \Phi_Y(\omega_y)$$

[3] We always have $\mathrm{COV}[X, Y] = \mathrm{COR}[X, Y] - m_x m_y$.

[4] $|\rho|$ is called strength of correlation, and $\mathrm{sgn}(\rho)$ is call direction of correlation.

[5] It is a 2D inverse Fourier transform. And we can get the pdf with

$$f_{X,Y}(x, y) = \frac{1}{(2\pi)^2} \iint_{-\infty}^{+\infty} \Phi_{X,Y}(\omega_x, \omega_y) e^{-j(\omega_x X + \omega_y Y)} \, \mathrm{d}\omega_x \, \mathrm{d}\omega_y$$

Theorem 18. ***(moment theorem)***

$$\mathbb{E}[X^{k_x} Y^{k_y}] = \frac{1}{j^{k_x+k_y}} \frac{\partial^{k_x}}{\partial \omega_x^{k_x}} \frac{\partial^{k_y}}{\partial \omega_y^{k_y}} \Phi_{X,Y}(\omega_x, \omega_y)\Big|_{\omega_x=\omega_y=0}$$

4.6 Gaussian random vectors

Definition 58. ***Jointly Gaussian r.v.*** *is defined with pdf*

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_x)^2}{\sigma_x^2} - \frac{2\rho(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2}\right]}$$

, where parameter $|\rho| < 1$*.*

Properties of Jointly Gaussian r.v.:

1. Jointly Gaussian $\Longrightarrow$ marginally Gaussian; while the converse does not hold.
2. $\rho = 0 \iff X \perp\!\!\!\perp Y$

5

Markov Chains