

A Digression on Probability Theory

Lecture Notes for CS 265

Delivered from January 28 through February 2, 2004

© Robert R. Snapp 2004

Motivation

It is impossible to know every factor that concerns the performance of computer networks:

- Network traffic is *bursty*.
- Unanticipated equipment failures occur.
- Economic, social, and political events can cause large (*nonstationary*) fluctuations in network traffic.
- Malicious events, such as viruses or denial-of-service attacks, can occur without warning.

When a phenomena cannot be described completely in deterministic terms, we describe it as being *random*, or *stochastic*. Nevertheless, such phenomena may exhibit a weak degree of regularity.

Probability theory is a useful tool for describing, analyzing, and predicting the regularity that exists in stochastic phenomena, including computer networking.

Elementary Probability Theory

A. N. Kolmogorov (1933) introduced a *probability space* as a triple (Ω, \mathcal{A}, P) :

Ω is the *set of elementary events*, or *sample space*.

\mathcal{A} is the *set of observable events*. An event A is a subset of Ω ; the *complement of A* is defined as $\bar{A} \triangleq \Omega \setminus A$.

\mathcal{A} is a σ -*algebra* of events:

- If $A \in \mathcal{A}$, then $\bar{A} \in \mathcal{A}$.
- If $A_i \in \mathcal{A}$, for $i = 1, 2, \dots$ (finite or countably infinite), then

$$\bigcup_i A_i \in \mathcal{A}.$$

- $\Omega \in \mathcal{A}$.

P is a *probability measure* that assigns a non-negative, real number to each element of \mathcal{A} , such that

- $P(\Omega) = 1$,
- If $A_i \in \mathcal{A}$, for $i = 1, 2, \dots$ (finite or countably infinite), are *mutually exclusive*, $A_i \cap A_j = \emptyset$ whenever $i \neq j$, then

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i). \quad (\text{complete additivity})$$

Example: Three consecutive coin tosses

The set of elementary events is finite:

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

The set of measurable events, \mathcal{A} has cardinality $|\mathcal{A}| = 2^{|\Omega|} = 256$.

Some elements of \mathcal{A} include

- Each elementary event: $\{HHH\}, \{HHT\}, \{HTH\}, \{HTT\}, \dots$
- Sequences with two or more heads: $\{HHH, HHT, HTH, THH\}$
- Sequences with one or fewer heads: $\{HTT, THT, TTH, TTT\}$
- Sequences with exactly one head $\{HTT, THT, TTH\}$
- Palindromic sequences $\{HHH, HTH, TTT, THT\}$
- Non-palindromic sequences $\{HHT, HTT, THH, TTH\}$
- First toss lands heads: $\{HHH, HTH, HHT, HTT\}$
- First toss lands tails: $\{THH, THT, TTH, TTT\}$
- Homogeneous sequences: $\{HHH, TTT\}$
- Heterogeneous sequences: $\{HHT, HTH, HTT, THH, THT, TTH\}$

Some Theorems

For a σ -algebra \mathcal{A}

1. If $A_i \in \mathcal{A}$, for $i = 1, 2, \dots, n$ then $\bigcap_{i=1}^n A_i \in \mathcal{A}$.

Proof: By DeMorgan's Law,

$$\bigcap_{i=1}^n A_i = \overline{\bigcup_{i=1}^n \overline{A_i}}.$$

By the definition of a σ -algebra, the right side of the above is in \mathcal{A} . ■

2. If $A, B \in \mathcal{A}$, then $A \setminus B \in \mathcal{A}$.

Proof: $A \setminus B = A \cap \overline{B}$. From the theorem above, $A \cap \overline{B} \in \mathcal{A}$. ■

Given a probability space (Ω, \mathcal{A}, P) , with $A, B \in \mathcal{A}$,

3. $P(\overline{A}) = 1 - P(A)$.
4. $P(\emptyset) = 0$.
5. $P(A) = 0 \not\Rightarrow A = \emptyset$
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Conditional Probability

Given a probability space (Ω, \mathcal{A}, P) , with $A, B \in \mathcal{A}$, such that $P(B) > 0$, the *conditional probability of A with respect to B* is defined as

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

Theorems

6. For $A_i \in \mathcal{A}$, for $i = 1, 2, \dots, n$

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1) P(A_2|A_1) P\left(A_3 \mid A_1 \cap A_2\right) \cdots P\left(A_n \mid \bigcap_{i=1}^{n-1} A_i\right)$$

7. $P(\Omega|A_i) = 1$.

8. If $A_i \in \mathcal{A}$, for $i = 1, 2, \dots$ (finite or countably infinite), are *mutually exclusive*, $A_i \cap A_j = \emptyset$ whenever $i \neq j$, and if $B \in \mathcal{A}$ with $P(B) > 0$, then

$$P\left(\bigcup_i A_i \mid B\right) = \sum_i P(A_i|B).$$

Bayes's Theorem

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

Proof: From the definition of conditional probability, or Theorem 6 with $n = 2$,

$$\begin{aligned} P(A \cap B) &= P(A|B) P(B), \\ &= P(B|A) P(A), \quad (\text{by symmetry}). \end{aligned}$$

Equating the two right sides of the above, yields the hypothesis. ■

Theorem (Total Probability): Let $A_i \in \mathcal{A}$, for $i = 1, 2, \dots, n$, be mutually exclusive, with $\bigcup_{i=1}^n A_i = \Omega$. Let $B \in \mathcal{A}$. Then,

$$P(B) = \sum_{i=1}^n P(B|A_i) P(A_i).$$

Proof: $B = \bigcup_{i=1}^n B \cap A_i$. Note that $(B \cap A_i) \cap (B \cap A_j) = \emptyset$, whenever $i \neq j$. ■

Combining the above, yields the more general theorem of Bayes:

$$P(A_j|B) = \frac{P(B|A_j) P(A_j)}{\sum_{i=1}^n P(B|A_i) P(A_i)}.$$

Independence

Two events $A, B \in \mathcal{A}$ are said to be *independent* if $P(A \cap B) = P(A)P(B)$.

Exercise: If A and B are independent, show that

$$P(\bar{A} \cap B) = P(\bar{A})P(B), \quad P(A \cap \bar{B}) = P(A)P(\bar{B}), \quad \text{and,} \quad P(\bar{A} \cap \bar{B}) = P(\bar{A})P(\bar{B}).$$

More generally, n events $A_i \in \mathcal{A}$, for $i = 1, 2, \dots, n$ are said to be *mutually independent* if all of the following equations are satisfied, for $m = 1, 2, 3, \dots, n$

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_m}) = P(A_{i_1})P(A_{i_2}) \cdot \dots \cdot P(A_{i_m}),$$

with, $1 \leq i_1 < i_2 < \dots < i_m \leq n$.

Example: $n = 3$ events, A , B , and C , are independent if and only if

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

$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

Bernoulli Trials

Consider the example of a coin being tossed n times.

Assume successive tosses are mutually independent, and let $p = P(H)$ and $q = 1 - p = P(T)$. (Note, $|\Omega| = 2^n$, and $|\mathcal{A}| = 2^{2^n}$.)

Then

$$b(i, n) = P(i \text{ heads in a sequence of } n \text{ tosses}) = \binom{n}{i} p^i q^{n-i}.$$

Note,

$$P(\Omega) = \sum_{i=0}^n b(i, n) = \sum_{i=0}^n \binom{n}{i} p^i q^{n-i} = (p + q)^n = 1.$$

Example:

$$\begin{aligned} P(1 \text{ or more heads in } n \text{ tosses}) &= 1 - b(0, n) \\ &= 1 - (1 - p)^n. \end{aligned}$$

Birthday Problem

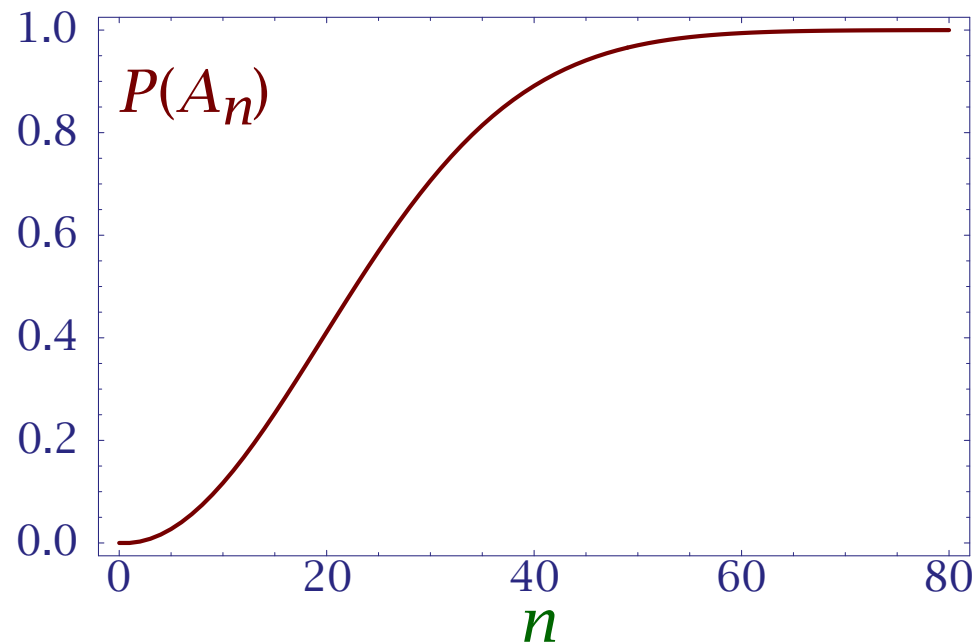
Let A_n denote the event that two or more people, in a group of n share the same birthday (neglecting leap years).

$$P(\overline{A_n}) = 1 \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{n-1}{365}\right).$$

Thus,

$$P(A_n) = 1 - P(\overline{A_n}) = 1 - 1 \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{n-1}{365}\right).$$

n	$P(A_n)$
10	0.116948
20	0.411438
30	0.706316
40	0.891232
50	0.970374
60	0.994123
70	0.999160
80	0.999914



Discrete Random Variables

An *integer-valued random variable* is a function $\mathbf{X} : \Omega \rightarrow S$, where $S \subset \mathbb{Z}$. (This definition is readily generalized to random variables that assume binary values, categorical values, and values from other discrete spaces.)

The discrete random variable \mathbf{X} is said to be *measurable* if

$$\{\omega \in \Omega : \mathbf{X}(\omega) = i\} \in \mathcal{A}, \forall i \in S.$$

The *probability distribution* of a discrete random variable $\mathbf{X} \in S$ is defined by

$$P_i = P(\mathbf{X} = i) \triangleq P\{\omega \in \Omega : \mathbf{X}(\omega) = i\}.$$

Common Discrete Distributions

Distribution	Parameters	P_i	S
Uniform	$n \in \mathbb{Z}^+$	$\frac{1}{n}$	$\{1, 2, \dots, n\}$
Binomial	$n \in \mathbb{Z}^+$ $p, q \geq 0,$ $p + q = 1$	$\binom{n}{i} p^i q^{n-i}$	$\{0, 1, \dots, n\}$
Geometric	$p, q \geq 0,$ $p + q = 1$	qp^i	\mathbb{Z}^+
Poission	$\lambda \in (0, +\infty)$	$e^{-\lambda} \frac{\lambda^i}{i!}$	\mathbb{Z}^+
Hypergeometric	$k, m, n \in \mathbb{Z}^+,$ $k \leq m + n$	$\frac{\binom{m}{i} \binom{n}{k-i}}{\binom{m+n}{k}}$	$\{0, 1, \dots, m\}$

Continuous Random Variables

A *real-valued random variable* is a function $\mathbf{X} : \Omega \rightarrow S \subset \mathbb{R}$,

A real-valued random variable is *measurable* if

$$\{\omega \in \Omega : \mathbf{X}(\omega) < x\} \in \mathcal{A}, \forall x \in S.$$

We define the *probability distribution* of the real-valued, random variable \mathbf{X} , as

$$F_{\mathbf{X}}(x) = P\{\mathbf{X} < x\} \triangleq P\{\omega \in \Omega : \mathbf{X}(\omega) < x\},$$

where the subscript of $F_{\mathbf{X}}$ is often omitted if no confusion arises.

Likewise, we define the *probability density* of $\mathbf{X} \in S$, as

$$f_{\mathbf{X}}(x) = \frac{d}{dx} F_{\mathbf{X}}(x)$$

at all points $x \in S$ where the derivative is defined.

Common Continuous Probability Densities

Distribution	Parameters	$f_X(x)$	S
Rectangular	$a, b \in \mathbb{R}$ $a < b$	$\frac{1}{b - a}$	$[a, b]$
Triangular	$a > 0$	$\frac{1}{a} \left(1 - \frac{ x }{a}\right)$	$[-a, a]$
Normal	$\mu \in \mathbb{R}$ $\sigma \in (0, +\infty)$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$	\mathbb{R}
Gamma	$\lambda, s \in (0, +\infty)$	$\left(\frac{x}{s}\right)^{\lambda-1} \frac{e^{-x/s}}{s\Gamma(\lambda)}$	$[0, \infty)$
Cauchy	$s \in (0, +\infty)$	$\frac{s}{\pi} \frac{1}{x^2 + s^2}$	\mathbb{R}
Beta	$a, b \in (0, +\infty)$	$\frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1 - x)^{b-1}$	$[0, 1]$

$$\Gamma(u) \triangleq \int_0^{\infty} t^{u-1} e^{-t} dt, \quad \Gamma(u + 1) = u\Gamma(u), \quad \Gamma(n + 1) = n!$$

Statistical Moments of Random Variables

Moment	Discrete	Continuous [*]
Mean	$E(\mathbf{X}) = \sum_{i \in \mathcal{S}} i P(\mathbf{X} = i)$	$E(\mathbf{X}) = \int_{\mathcal{S}} x f_{\mathbf{X}}(x) dx$
k -th moment	$E(\mathbf{X}^k) = \sum_{i \in \mathcal{S}} i^k P(\mathbf{X} = i)$	$E(\mathbf{X}^k) = \int_{\mathcal{S}} x^k f_{\mathbf{X}}(x) dx$
Variance	$\text{Var}(\mathbf{X}) \triangleq E\left((\mathbf{X} - E(\mathbf{X}))^2\right) = E(\mathbf{X}^2) - E(\mathbf{X})^2$	

$$E(a\mathbf{X} + b\mathbf{Y}) = aE(\mathbf{X}) + bE(\mathbf{Y}), \quad \text{Var}(a\mathbf{X}) = a^2 \text{Var}(\mathbf{X})$$

* If the density $f_{\mathbf{X}}(x)$ is undefined, these integrals can be evaluated as Stieltjes integrals, i.e.,

$$E(g(\mathbf{X})) = \int_{\mathcal{S}} g(x) dF_{\mathbf{X}}(x) \triangleq \lim_{\delta \rightarrow 0} \sum_{i=-\infty}^{\infty} \sup_{i\delta < x \leq (i+1)\delta} g(x) (F_{\mathbf{X}}((i+1)\delta) - F_{\mathbf{X}}(i\delta)),$$

where we assume the same value is obtained if **sup** is replaced by **inf**.

Example: Moments of a Binomial Random Variable

Let $\mathbf{X} \sim \text{Binomial}(n, p)$, i.e.,

$$P(\mathbf{X} = i) = b(i, n) = \binom{n}{i} p^i q^{n-i}$$

for $i = 0, 1, \dots, n$, where $q = 1 - p$.

$$\begin{aligned} E(\mathbf{X}) &= \sum_{i=0}^n i \cdot P(\mathbf{X} = i) = \sum_{i=0}^n i \binom{n}{i} p^i q^{n-i} = p \frac{\partial}{\partial p} \sum_{i=0}^n \binom{n}{i} p^i q^{n-i} \\ &= p \frac{\partial}{\partial p} (p + q)^n = np(p + q)^{n-1} = np \end{aligned}$$

$$\begin{aligned} \text{Var}(\mathbf{X}) &= E(\mathbf{X}^2) - E(\mathbf{X})^2 = E(\mathbf{X}(\mathbf{X} - 1)) + E(\mathbf{X}) - E(\mathbf{X})^2 \\ E(\mathbf{X}(\mathbf{X} - 1)) &= \sum_{i=0}^n i(i-1) \binom{n}{i} p^i q^{n-i} = p^2 \frac{\partial^2}{\partial p^2} \sum_{i=0}^n \binom{n}{i} p^i q^{n-i} \\ &= p^2 \frac{\partial^2}{\partial p^2} (p + q)^n = n(n-1)p^2(p + q)^{n-2} = n(n-1)p^2 \\ \text{Var}(\mathbf{X}) &= n(n-1)p^2 + np - (np)^2 = np(1-p) = npq \end{aligned}$$

Functions of a single random variable

Let $\mathbf{X} : \Omega \rightarrow S \subset \mathbb{R}$, and let $g : S \rightarrow S' \subset \mathbb{R}$ be sufficiently regular, so that

$$\{\omega \in \Omega : g(\mathbf{X}(\omega)) < y\} \in \mathcal{A}, \forall y \in S'.$$

Then $\mathbf{Y} = g(\mathbf{X})$ is a measurable random variable with probability distribution

$$F_{\mathbf{Y}}(y) = P(\mathbf{Y} < y) = P\{\omega \in \Omega : g(\mathbf{X}) < y\}.$$

For example, if g is also a one-to-one, *increasing* function, then

$$F_{\mathbf{Y}}(y) = F_{\mathbf{X}}(g^{-1}(y)), \quad \text{whence, } f_{\mathbf{Y}}(y) = f_{\mathbf{X}}(g^{-1}(y)) \cdot \frac{dg^{-1}}{dy}.$$

If g is also a one-to-one, *decreasing* function, then

$$F_{\mathbf{Y}}(y) = 1 - F_{\mathbf{X}}(g^{-1}(y)),$$

whence,

$$f_{\mathbf{Y}}(y) = -f_{\mathbf{X}}(g^{-1}(y)) \cdot \frac{dg^{-1}}{dy} = f_{\mathbf{X}}(g^{-1}(y)) \cdot \left| \frac{dg^{-1}}{dy} \right|.$$

Multiple Random Variables

Sometimes problems will involve more than one random quantity, e.g.,

$$\mathbf{X}_i : \Omega \rightarrow S_i \subset \mathbb{R}, \quad \text{for } i = 1, 2, \dots, n.$$

These variables are measurable, if $\forall x_1 \in S_1, \dots, \forall x_n \in S_n$,

$$\{\omega \in \Omega : \mathbf{X}_1(\omega) < x_1, \dots, \mathbf{X}_n(\omega) < x_n\} \in \mathcal{A}.$$

We define the *joint probability distribution* as

$$\begin{aligned} F_{\mathbf{X}_1, \dots, \mathbf{X}_n}(x_1, \dots, x_n) &\triangleq P(\mathbf{X}_1 < x_1, \dots, \mathbf{X}_n < x_n) \\ &= P\{\omega \in \Omega : \mathbf{X}_1(\omega) < x_1, \dots, \mathbf{X}_n(\omega) < x_n\}. \end{aligned}$$

The *joint probability density* is defined as

$$f_{\mathbf{X}_1, \dots, \mathbf{X}_n}(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{\mathbf{X}_1, \dots, \mathbf{X}_n}(x_1, \dots, x_n).$$

The random variables $\mathbf{X}_1, \dots, \mathbf{X}_n$ are said to be *independent*, if

$$F_{\mathbf{X}_1, \dots, \mathbf{X}_n}(x_1, \dots, x_n) = F_{\mathbf{X}_1}(x_1) \cdots F_{\mathbf{X}_n}(x_n).$$

Functions of Two or More Random Variables

Let $\mathbf{X} \in \mathbb{R}$ and $\mathbf{Y} \in \mathbb{R}$ be independent RVs with distributions $F_{\mathbf{X}}(x)$ and $F_{\mathbf{Y}}(y)$.

(a) Find the distribution and density of the sum $\mathbf{S} = \mathbf{X} + \mathbf{Y}$:

$$\begin{aligned} F_{\mathbf{S}}(s) &= P(\mathbf{X} + \mathbf{Y} < s) \\ &= \int_{\mathbb{R}} F_{\mathbf{Y}}(s - x) dF_{\mathbf{X}}(x) = \int_{\mathbb{R}} F_{\mathbf{Y}}(s - x) f_{\mathbf{X}}(x) dx \\ &= \int_{\mathbb{R}} F_{\mathbf{X}}(s - y) dF_{\mathbf{Y}}(y) = \int_{\mathbb{R}} F_{\mathbf{X}}(s - y) f_{\mathbf{Y}}(y) dy. \end{aligned}$$

The probability density of \mathbf{S} is thus

$$f_{\mathbf{S}}(s) = \frac{d}{ds} F_{\mathbf{S}}(s) = \int_{\mathbb{R}} f_{\mathbf{Y}}(s - x) f_{\mathbf{X}}(x) dx = \int_{\mathbb{R}} f_{\mathbf{X}}(s - y) f_{\mathbf{Y}}(y) dy.$$

(b) Find the distribution and density of the maximum $\mathbf{U} = \max(\mathbf{X}, \mathbf{Y})$:

For the distribution,

$$F_{\mathbf{U}}(u) = P(\mathbf{X} < u, \mathbf{Y} < u) = F_{\mathbf{X}}(u)F_{\mathbf{Y}}(u).$$

For the density,

$$f_{\mathbf{U}}(u) = \frac{d}{du} F_{\mathbf{U}}(u) = f_{\mathbf{X}}(u)F_{\mathbf{Y}}(u) + F_{\mathbf{X}}(u)f_{\mathbf{Y}}(u).$$