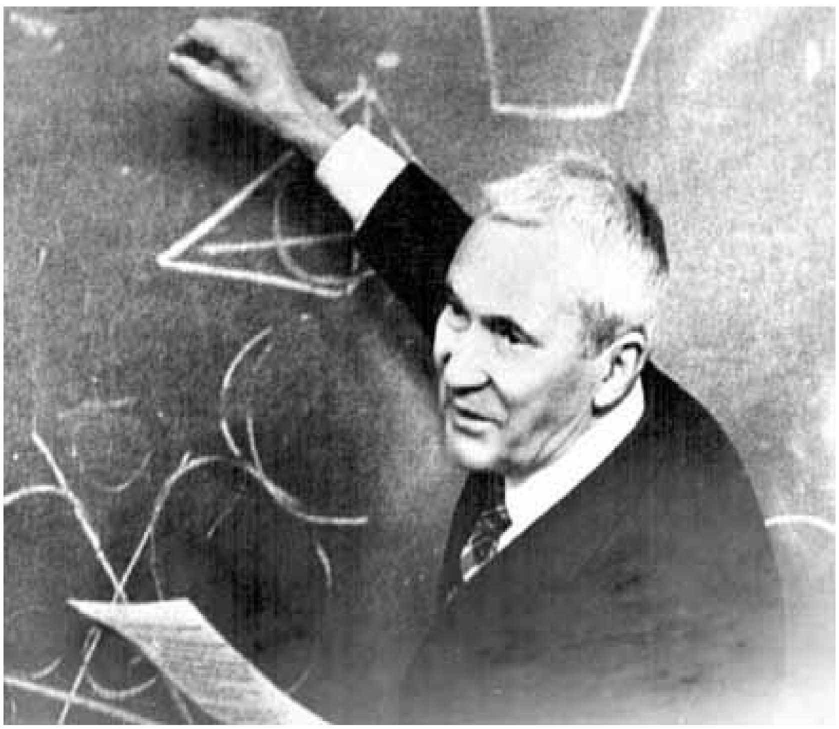
Probability Review

Kolmogorov's Axioms of Probability



Andrey Kolmogorov (1903-1987)

One of the fundamental tasks of simulation is to sample elements from a set S according to some probability distribution $p(x)$, where $p(x)$ represents the likelihood of sampling $x \in S$. For this reason S is referred to as a **sample space**. Moreover, if S is finite or countably infinite, then it is possible to define such a **probability distributiion** $p(x)$ for which

1. $0 \leq p(x) \leq 1$, for all $x \in S$, and
2. $\sum_{x \in S} p(x) = 1$.

Example 1. Let S represent the set of possible values that can apear on the faces of a pair of dice. For example, the if the face of the first die D_1 shows 1, while the face of the second die D_2 shows 5, then this can be represented as $(1,5) \in S$. Thus, $|S| = 6 \times 6 = 36$ and $p((a,b)) = 1/36$, $1 \le a, b \le 6$, defines a probability distribution over S.

In later lectures we examine efficient sampling algorithms; i.e. algorithms for sampling from S in accordance with some probability distribution over S. Of course, sampling is occurring all the time in nature. For example, to obtain samples for set S from Example 1, we may visit a craps table at a Las Vegas casino, and record the outcomes of a gambler's dice rolls. However, what makes computer simulation so powerful is that a modern computer can generate more samples in one second than all the craps tables combined in one year. This allows computer simulations to obtain accurate estimates of statistics related to S (including its size, mean, and variance), as well as the likelihood of events associated with S.

An **event** E of a sample space S is a subset of S for which the likelihood/probability of E, denoted $P(E)$, can be measured. For example, if S is finite or countably infinite with probability distribution $p(x)$, then *every* subset E of S is an event, with

$$P(E) = \sum_{x \in E} p(x).$$

Example 2. Consider the sample space S from Example 1. Then the act of rolling a seven with the dice is an event of S, since it may be represented with the set

$$E = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\} \subset S.$$

Moreover, using the distribution $p(x) = 1/36$, it follows that $P(E) = 6/36 = 1/6$. What is the probability of rolling an even number?

In the case when S is uncountable, problems arise when attempting to measure the likelihood of every subset of S. So instead we only require that the set $\mathcal{E}$ of events forms what is called a **σ-algebra**, meaning that i) $S \in \mathcal{E}$, ii) $\emptyset \in \mathcal{E}$, iii) if $E \in \mathcal{E}$, then $\overline{E} = S - E \in \mathcal{E}$, and iv) $\mathcal{E}$ is closed under countable unions and intersections. In other words, if $E_1, E_2, \ldots,$ are all events, then so are $\bigcup\limits_{i=1}^{\infty} E_i$ and $\bigcap\limits_{i=1}^{\infty} E_i$.

Example 3. Let $\mathcal{R}$ denote the set of real numbers, and suppose that $\mathcal{E}$ contains all intervals of the form (a, b), $[a, b]$, $(a, b]$, $[a, b)$, where $a < b$ are real numbers. Furthermore, assume that $\mathcal{E}$ is closed under countable unions and intersections, and complements. Verify that i) $\mathcal{E}$ is a σ-algebra with respect to $\mathcal{R}$, and that ii) $\mathcal{E}$ contains all singleton sets of the form $\{a\}$, where a is a real number. Note: this σ-algebra is commonly referred to as the **Borel algebra**.

Note that σ-algebras are usually defined in a manner similar to Example 3. In other words, first begin with a base family $\mathcal{B}$ of sets (e.g. finite intervals) that you desire to be in the algebra, then define $\mathcal{E}$ as the smallest σ-algebra containing $\mathcal{B}$. This is referred to as the **σ-algebra generated by $\mathcal{B}$**.

Now suppose $\mathcal{E}$ is a σ-algebra with respect to sample space S, and a probability $P(E)$ has been defined for each $E \in \mathcal{E}$. Then Kolmogorov's axioms of probability place the following restrictions on P.

1. $P(S) = 1$
2. For pairwise disjoint subsets $E_1, E_2, \ldots,$

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

Whenever a function $P : \mathcal{E} \to [0, 1]$ satisfies the above axioms, then P is called a **probability measure** over $\mathcal{E}$, and $(S, \mathcal{E}, P)$ is called a **probability measure space**.

Proposition 1. Let $(S, \mathcal{E}, P)$ be a probability measure space. Then the following statements hold for arbitrary $A, B \in \mathcal{E}$.

1. $P(\emptyset) = 0$

2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

3. $P(\overline{A}) = 1 - P(A)$

Proof of Propositiion 1. We prove the first two statements and leave the third as an exercise. By Axiom 2, $P(A) = P(A \cup \emptyset) = P(A) + P(\emptyset)$ which implies $P(\emptyset) = 0$.

For proving the second statement, first it is an exercise to show that a σ-algebra is closed under set difference. In other words, if $A, B \in \mathcal{E}$, then $A - B \in \mathcal{E}$. Then by Axiom 2 we have $P(A) = P(A - B) + P(A \cap B)$, which implies $P(A - B) = P(A) - P(A \cap B)$. Similarly, $P(B - A) = P(B) - P(A \cap B)$. Also by Axiom 2,

$$P(A \cup B) = P(A - B) + P(B - A) + P(A \cap B) = (P(A) - P(A \cap B)) + (P(B) - P(A \cap B)) + P(A \cap B) =$$

$$P(A) + P(B) - P(A \cap B).$$

Conditional Probability

Given a probability measure space $(S, \mathcal{E}, P)$, let A and B be two events, with $P(B) > 0$. Then $P(A|B)$ denotes the **probability of A on condition B (has been observed)** and is defined as

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

where $P(A, B)$ is a convenient way of writing $P(A \cap B)$. Intuitively, if we think of $P(B)$ as an area, then $P(A|B)$ is measuring the fraction of that area that overlaps with event A, and hence represents the likelihood of A being observed given that B has already been observed.

Example 4. Suppose you learn that one of the children from a two-child family is a girl. What is the probability that the second child is also a girl? Assume the child combinations boy-boy, boy-girl, girl-boy, girl-girl are all equally likely.

Example 5. Three coins are tossed independently. Coin i, $i = 1, 2, 3$, has probability $i/4$ of landing heads. What is the probability that Coin 1 has landed heads, on condition that two out of the three coins landed heads? one out of three? zero out of three? three out of three?

The **Law of Total Probability** shows how conditional probabilities can seem useful for computing unconditional ones. Let A be an event whose probability is to be computed, and let $E_1, E_2, \ldots, E_n$ be pairwise disjoint events for which $P(\overset{n}{\underset{i=1}{\cup}} E_i) = 1$ (i.e., exactly one of the E_i will be observed). Then

$$P(A) = \sum_{i=1}^{n} P(A|E_i)P(E_i).$$

The proof is left as an exercise.

Example 6a. Suppose a bag contains three coins. One coin is a two-headed coin, the other is a fair coin, while the third has only a probability of 0.25 of landing heads. If one of the coins is randomly drawn from the bag and tossed, what is the probability that a heads will be observed?

Two events A and B are called **independent** iff $P(A, B) = P(A)P(B)$. Moreover, if one of the events, say B, has $P(B) > 0$, then A and B are independent iff $P(A|B) = P(A)$. In other words, observing B does not change the likelihood of observing A, and vice versa (assuming $P(A) > 0$).

Example 6b. Consider the sample space and probability distribution from Example 1. Show that the value observed on the first die is independent of that observed on the second.

Random Variables

When a sample space S consists of numbers (such as integers or real numbers) then it is referred to as a **random variable**, while its set of values is referred to as its **domain**. Random variables will usually be denoted with higher-end capital letters, such as S, T, X, Y, etc.. A random variable X typically falls into one of the following three categories.

Finite $|X|$ is finite

Discrete $|X|$ is finite or countably infinite

Continuous X consists of one or more intervals of the real number line

The σ-algebra associated with a random variable X is usually generated by events of the form $X \leq a$, for some real number a. It is an exercise to show that, if X is finite or discrete, then this algebra consists of all subsets of X. We refer to this algebra as the **discrete algebra**. Moreover, if $X = \mathcal{R}$, then this algebra is equal to the Borel algebra from Example 3. Henceforth we assume that all random variables are either discrete or continuous, and have either the discrete or Borel algebra.

As before, a finite or discrete random variable X has a corresponding probability distribution function $p(x)$ that assigns a probability to each element $x \in X$. However, if X is continuous, then the probability of a single element x will usually be zero, and so, instead of a point-to-point probability distribution, what is provided is a **probability density function** $f(x)$ that is used to compute the (usually nonzero) probabilities of interval events. Moreover, just as a probability distribution must sum to unity, in the case of a density function $f(x)$, we require that

$$\int_X f(x)dx = 1.$$

Thus, to compute the probability of an interval $[a, b]$, one evaluates the integral $\int_a^b f(x)dx$.

Example 7. Verify that $f(x) = 3e^{-3x}$ is a probability density function for $X = [0, \infty)$, and use it to compute $P(X \leq 1)$.

We note in passing that random variables $X_1, \ldots, X_n$, $n \geq 2$, can be combined to form a **random vector** $Z = (X_1, \ldots, X_n)$, where the domain of Z is now a subset of $\mathcal{R}^n$. Each X_i, $i = 1, \ldots, n$, is called a **component variable**. Furthermore, when all the component variables are discrete, then the associated **joint probability distribution** is a function $p(x_1, \ldots, x_n)$ that sums to unity. On the other hand, if one or more of the component variables is continuous, then associated with Z is a **joint density function** $f(x_1, \ldots, x_n)$ for which

$$\int_{\mathcal{R}^n} f(x_1, \ldots, x_n)dx_1 \cdots dx_n = 1.$$

Now events are subsets E of $\mathcal{R}^n$, where $P(E)$ is obtained by integrating (respectively, summing) the joint density (respectively, distribution) over region E.

Cumulative distribution function

Associated with each random variable X is a **cumulative distribution function (cdf)** $F(x)$ with the following properties.

1. $F(x)$ is defined for every real number $x \in \mathcal{R}$.
2. $F(x) = P(X \leq x)$.
3. F is nondecreasing, i.e. if $a \leq b$ then $F(a) \leq F(b)$.
4. $\lim_{x\to\infty} F(x) = 1$
5. $\lim_{x\to-\infty} F(x) = 0$
6. $F(x)$ is right-continous; i.e. $\lim_{x\to a^+} F(x) = F(a)$
7. Formulas for computing $F(x)$
   (a) **discrete.** $F(x) = \sum\limits_{t \leq x} p(t)$
   (b) **continuous.** $F(x) = \int_{-\infty}^{x} f(t)dt$

Example 8. Compute and graph the cdf of i) the random variable X having domain $\{1, 2, 3\}$ and probability distribution $p(1) = p(2) = 0.25$, $p(3) = 0.5$, and ii) the random variable from Example 7.

Two random variables X and Y are said to be **independent** iff the events $X \le a$ and $Y \le b$, are independent, for arbitrary $a, b \in \mathcal{R}$. The following fact is stated without proof.

Proposition 2. Random variables X and Y are independent iff the joint probability distribution (respectively, density) function of $Z = (X, Y)$ has the form $p(x, y) = p(x)p(y)$ (respectively, $f(x, y) = f(x)f(y)$).

Example 9. Let X and Y denote the outcomes of two distinct die. Let $Z = X + Y$. Show that X and Y are independent, but that X and Z are not.

Statistics of a random variable

Associated with a random variable X is a collection of **prior statistics** that provide information about the distributional tendencies that one encounters when observing samples of X. These can be computed using the density or distribution function associated with X *prior* to observing any samples.

Expectation the average value assumed by X

1. also referred to as the **average** or **mean**
2. **discrete.** $E[X] = \sum_x xp(x)$
3. **continuous.** $E[X] = \int_{-\infty}^{\infty} xf(x)dx$

Variance measures the average (square of the) distance from the mean

1. **discrete.** $\text{Var}(X) = \sum_x (x - E[X])^2 p(x)$
2. **continuous.** $\text{Var}(X) = \int_{-\infty}^{\infty} (x - E[X])^2 f(x)dx$

Standard Deviation measures the average distance from the mean

$$\sigma_X = \sqrt{\text{Var}(X)}$$

Covariance Measures the degree of dependence between two variables X and Y

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

n **th Moment** $E[X^n]$

Mode the element(s) of X with the highest probability or density value value

Proposition 3. The following identifies hold.

1. **Linearity of expectation**
$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$$

2. **Variance identity**
$$\mathrm{Var}(X) = E[X^2] - E^2[X]$$

3. If X and Y are independent, then $E[XY] = E[X]E[Y]$.

4. **Linearity of variance under independence** If $X_1, \ldots, X_n$ are independent, then
$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

5. **Covariance of independent variables** If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$.

Proof of Proposition 3. We prove Identities 1 and 2 and leave the other three as exercises.

For simplicity, we prove linearity of expectation for $n = 2$ and using finite random variables $X = \{x_1, \ldots, x_m\}$ and $Y = \{y_1, \ldots, y_n\}$. Proving it for arbitrary values of n can be accomplished with mathematical induction, while proving it for continuous variables can be accomplished uing the theory of integration.

Then
$$E[X+Y] = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_i + y_j)P(X = x_i, Y = y_j) =$$
$$\sum_{i=1}^{m}\sum_{j=1}^{n} x_i P(X = x_i, Y = y_j) + \sum_{i=1}^{m}\sum_{j=1}^{n} y_j P(X = x_i, Y = y_j) =$$
$$\sum_{i=1}^{m} x_i (\sum_{j=1}^{n} P(X = x_i, Y = y_j)) + \sum_{j=1}^{n} y_j (\sum_{i=1}^{m} P(X = x_i, Y = y_j)) =$$
$$\sum_{i=1}^{m} x_i P(X = x_i) + \sum_{j=1}^{n} y_j P(Y = y_j) =$$
$$E[X] + E[Y].$$

To prove Identity 2, first notice that $\mathrm{Var}(X)$ may be defined as $\mathrm{Var}(X) = E[(X - E[X])^2]$, which is equivalent to the above definition. Then after squaring the expression and using linearity of expectation,
$$E[(X - E[X])^2] = E[X^2 - 2XE[X] + E^2[X])] =$$
$$E[X^2] - E[2XE[X]] + E[E^2[X]] = E[X^2] - 2E[X]E[X] + E^2[X] = E[X^2] - E^2[X].$$
Note that the second-to-last equality uses the fact that $E[aX] = aE[X]$, for arbitrary constant a, where, in this case, $a = 2E[X]$.

Chebyshev's Inequality and the Law of Large Numbers

Proposition 4: Markov's Inequality. Let X be a nonnegative random variable. Then for any $a > 0$,

$$P(X \geq a) \leq \frac{E[X]}{a}.$$

Proof of Proposition 4. Define random variable $Y = \{0, a\}$, where $Y = a$ if $X \geq a$, and $Y = 0$ otherwise. Then, since X is nonnegative, we must have $X \geq Y$ with probabilty one. Therefore,

$$E[X] \geq E[Y] = aP(X \geq a),$$

which yields the desired inequality.

Corollary: Chebyshev's Inequality. If X is random variable with mean μ and standard deviation σ, then

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Proof of Corollary. Let $Z = \frac{(X-\mu)^2}{\sigma^2}$. Then Z is a nonnegative random variable with

$$E[Z] = E\left[\frac{(X-\mu)^2}{\sigma^2}\right] = \frac{1}{\sigma^2}E[(X-\mu)^2] = 1.$$

Then, by Markov's inequality,

$$P\left(\frac{(X-\mu)^2}{\sigma^2} \geq k^2\right) \leq \frac{1}{k^2}.$$

But the event

$$\frac{(X-\mu)^2}{\sigma^2} \geq k^2$$

is equivalent to the event

$$(X-\mu)^2 \geq (k\sigma)^2,$$

which in turn is equilvant to

$$|X - \mu| \geq k\sigma.$$

Theorem 1: Weak Law of Large Numbers. Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with mean μ, and standard deviation σ. Then, for any $\epsilon > 0$,

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) \to 0 \text{ as } n \to \infty.$$

Proof of Theorem 1: By linearity of expectation,

$$E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}(n\mu) = \mu.$$

Also, by independence,

$$\mathrm{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}.$$

Thus, Chebyshev's inequality yields

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \frac{k\sigma}{\sqrt{n}}\right) \leq \frac{1}{k^2}.$$

Now let $\epsilon > 0$ be arbitrary. Since k is also arbitrary in Chebyshev's inequality, we may set $k = \frac{\epsilon\sqrt{n}}{\sigma}$. Substituting this into the above inequality yields

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Finally, notice that the right side of this inequality is approaching zero as $n \to \infty$.

The Weak Law of Large Numbers has important consequences for simulation. In words, it implies that, the more independent samples we take of a random variable X, the more the average of these samples will approach $E[X]$ with a probability that increases with the number of samples.

Discrete Distributions

Bernoulli $B(p)$ 1. $X = \{0, 1\}$

2. $P(X = 1) = p$, $P(X = 0) = 1 - p$
3. $E[X] = p$, $\mathrm{Var}(X) = p(1 - p)$

Binomial $B(n, p)$ 1. $X = \{0, 1, \ldots, n\}$

2. $P(X = k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1 - p)^{n-k}$
3. $E[X] = np$, $\mathrm{Var}(X) = np(1 - p)$
4. **Practical Importance:** if one performs n independent experiments, each with probability p of success, then the number of successes follows the binomial distribution.

Geometric $G(p)$ 1. $X = \{1, 2, \ldots\}$

2. $P(X = k) = (1 - p)^{k-1} p$
3. $E[X] = \frac{1}{p}$, $\mathrm{Var}(X) = \frac{1-p}{p^2}$
4. **Practical Importance:** if one continually performs independent experiments (each with probability p of success) until a success occurs, then the number of experiments needed follows the geometric distribution.

Poisson Poisson(λ) 1. $X = \{0, 1, \ldots\}$

2. $P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$
3. $E[X] = \lambda$, $\mathrm{Var}(X) = \lambda$
4. **Practical Importance:** If the interarrival times of customers entering a system follows an exponential distribution with rate λ, then the number of arrivals per unit time is distrubted as Poisson(λ).
5. **Practical Importance:** Poisson(λ) can be used to approximate the Binomial distribution for large n and small p for which $\lambda = np$.

Negative Binomial NB(r, p) 1. $X = \{0, 1, \ldots\}$

2. $\mathrm{P}(X = k) = p^r \left(\begin{array}{c} r + k - 1 \\ k \end{array} \right) (1 - p)^k$
3. $E[X] = \frac{r(1-p)}{p}$, $\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$
4. **Practical Importance:** if one continually performs independent experiments (each with probability p of success) until r successes occur, then the number of experiments needed beyond r follows the negative binomial distribution.

Hypergeometric HG(m, n, r) 1. $X = \{0, 1, \ldots, r\}$

2. $\mathrm{P}(X = k) = \dfrac{\binom{m}{k}\binom{n}{r-k}}{\binom{m+n}{r}}$

3. $E[X] = \frac{mr}{m+n}$, $\mathrm{Var}(X) = \frac{mnr(m+n-r)}{(m+n)^2(m+n-1)}$

4. **Practical Importance:** given an urn with m blue balls and n red balls, it provides the probability of selecting k blue balls when selecting a total of r balls.

We use the notation $X \sim \mathcal{D}$ to denote that random variable X is a type-$\mathcal{D}$ random variable. For example, $X \sim \mathrm{Be}(p)$ means that X is a Bernoulli random variable.

Example 10. An unfair coin has a probability of 0.7 of landing heads. What is the probability of the unfair coin being independently tossed 10 times, and landing tails at most 2 times? How many tosses should we expect before 10 heads are witnessed?

Example 11. In a production process designed for canned tuna, it has been deemed acceptable for some cans to be "defective" so long as they constitute less than 0.1% of all produced cans. Suppose that 1,000 cans have been randomly inspected throughout the week and 10 of those cans were found to be defective. How likely is it to have 10 or more defects assuming that the probability of producing a defective can is .001?

Example 12. If a random hash function is hashing a key into a hash table that is 80% filled to capacity, what is the probability that it will take more than 10 attempts to find an open cell to store the data? Assume that each cell is equally likely to be selected at each attempt.

Example 13. If the number of hits per minute for a website is a Poisson random variable with parameter $\lambda = 3$, what is the probability that no hits occur within the next minute? What is the probability that 10 or more hits occur within the next 3 minutes?

Some Common Continuous Distributions

Uniform $U(a,b)$ 1. $X = [a,b]$

2. **density:**
$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$
3. **CDF:**
$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a}, & a \le x < b \\ 1 & x \ge b \end{cases}$$
4. $E[X] = \frac{a+b}{2}$, $\text{Var}(X) = \frac{(b-a)^2}{12}$
5. **Practical Importance:** used to model a variable use value falls between a and b, and is equally likely to take on any value between a and b.

Exponential $E(\lambda)$ 1. $X = [0, \infty]$

2. **density:** $f(x) = \lambda e^{-\lambda x}$
3. **CDF:** $F(x) = 1 - e^{-\lambda x}$
4. $E[X] = \frac{1}{\lambda}$, $\text{Var}(X) = \frac{1}{\lambda^2}$
5. **Practical Importance:** quite often used to model the interarrival times for a population of customers who are arriving independently, and with a rate of λ.

Normal $N(\mu, \sigma^2)$ 1. $X = \mathcal{R}$

2. **density:**
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
3. **CDF:** a normal distribution does not possess a CDF with a closed form. We let $\Phi(x)$ denote the CDF for the normal distribution $N(0,1)$.
4. $E[X] = \mu$, $\text{Var}(X) = \sigma^2$
5. **Practical Importance:** the Normal distribution is the distribution over $\mathcal{R}$ of maximum entropy on condition of having a given mean μ and variance σ^2.
6. **Standard Normal.** $N(0,1)$ is called the **standard normal distribution**. It can be shown that, if $X \sim N(\mu, \sigma^2)$, then $(X-\mu)/\sigma$ is distributed as $N(0,1)$.
7. **Central Limit Theorem.** Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with mean μ, and standard deviation σ. Furthermore, let
$$Y = \frac{X_1 + \cdots + X_n}{n}.$$
Then $E[Y] = \mu$ and $\text{Var}(Y) = \sigma^2/n$ (see proof of Weak Law of Large Numbers). Hence $Z = (Y-\mu)/\sigma/\sqrt{n}$ has zero mean and unit variance. Moreover, as n increases, the distribution of Z approaches that of $N(0,1)$. Therefore, for large values of n, the CDF $\Phi(x)$ may be used to approximate the probability of events pertaining to Z.

Weibull $W(\alpha, \beta, \nu)$ 1. $X = [\nu, \infty]$

2. **density:**

$$f(x) = \frac{\beta}{\alpha}(\frac{x-\nu}{\alpha})^{\beta-1}\exp[-(\frac{x-\nu}{\alpha})^{\beta}]$$

3. $E[X] = \nu + \alpha\Gamma(\frac{1}{\beta}+1)$, $\text{Var}(X) = \alpha^2[\Gamma(\frac{2}{\beta}+1) - \Gamma^2(\frac{1}{\beta}+1)]$
4. **Practical Importance:** is often good at modeling the time to failure for a system component.

Triangular $T(a, b, c)$ 1. $X = [a, c]$

2. **density:**

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)}, & a \le x \le b \\ \frac{2(c-x)}{(c-b)(c-a)}, & b < x \le c \end{cases}$$

where $a < b < c$

3. **CDF:**

$$F(x) = \begin{cases} 0 & x < a \\ \frac{(x-a)^2}{(b-a)(c-a)}, & a < x \le b \\ 1 - \frac{(c-x)^2}{(c-b)(c-a)} & b < x \le c \\ 1 & x > c \end{cases}$$

4. $E[X] = \frac{a+b+c}{3}$
5.

$$\text{Var}(X) = \frac{2}{(b-a)(c-a)}[\frac{1}{4}b^4 - \frac{ab^3}{3} + \frac{7}{12}a^4] + \frac{2}{(c-b)(c-a)}[\frac{1}{12}c^4 - \frac{b^4}{4} + \frac{cb^3}{3}] - \frac{(a+b+c)^2}{9}$$

6. **Practical Importance:** similar to the normal distribution, but values lie within a finite interval.

Log Normal LN(μ, σ^2) 1. $X = (0, \infty]$

2. **density:**

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}},$$

3. $E[X] = e^{\mu + \frac{\sigma^2}{2}}$, $\text{Var}(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$
4. **Practical Importance:** for situations when the data shows a normal tendency after it has been logarithmically scaled.

Gamma Ga(λ, n) 1. $X = [0, \infty]$

2. **density:**

$$f(x) = \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x},$$

where $\lambda, n > 0$, and $\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}dx$

3. $E[X] = \frac{n}{\lambda}$, $\text{Var}(X) = \frac{n}{\lambda^2}$

4. **Practical Importance:** parameters λ and n make the gamma distribution versatile for modeling miscellaneous random variables. When n is a positive integer, then X represents the sum of n independent exponential random variables having common rate λ. Yields the Chi-Square distribution when $\lambda = 1/2$ and $n = \frac{r}{2}$, where positive integer r is the number of degrees of freedom.

Beta Be(α, β)

1. $X = (0, 1)$
2. **density:**
$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$
where $\Gamma(\beta) = \int_0^\infty x^{\beta-1}e^{-x} = (\beta-1)!$ if β is a positive integer.
3. $E[X] = \frac{\alpha}{\alpha+\beta}$, $\mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
4. **Practical Importance:** often used as a distribution for determining the value p that is used for a particular binomial random variable whose p parameter is unknown.

Cauchy $C(\mu, \sigma^2)$

1. $X = (-\infty, \infty)$
2. **density:**
$$f(x) = \frac{\sigma}{\pi} \frac{1}{\sigma^2 + (x-\mu)^2}$$
3. $E[X]$ does not exist, $\mathrm{Var}(X)$ does not exist
4. **CDF:** $F(x) = \frac{\tan^{-1}(\frac{x-\mu}{\sigma})}{\pi} + \frac{1}{2}$

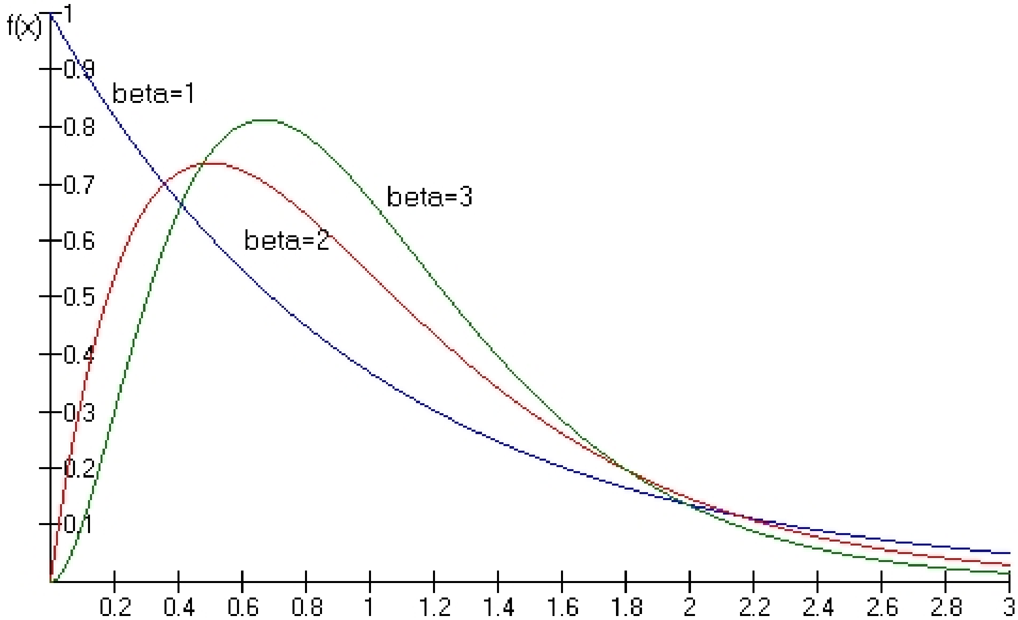


Figure 1. Some gamma density functions with $\lambda = 1$ and $n = \text{beta} = 1, 2, 3$.

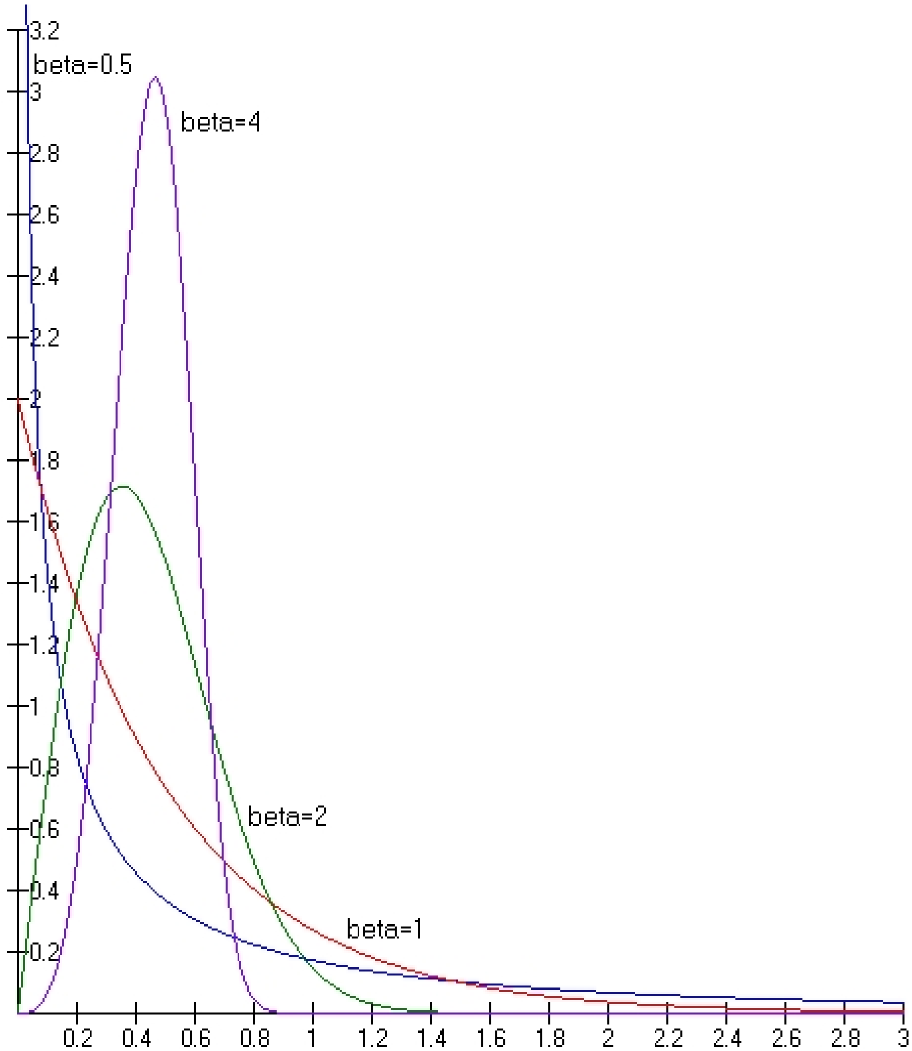


Figure 2. Some Weibull density functions with $\theta = 0$, $\alpha = 0.5$, and $\beta = 0.5, 1, 2, 4$.

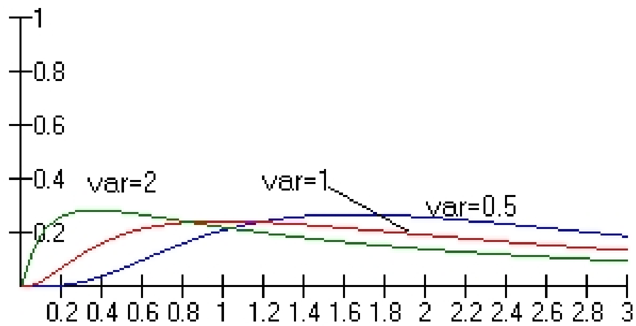


Figure 3. Some Weibull density functions with $\theta = 0$, $\alpha = 0.5$, and $\beta = 0.5, 1, 2, 4$.

Example 14. If $X \sim U(0, 10)$, compute $Pr(X \leq 3)$, $Pr(X \geq 7)$, and $Pr(1 \leq X \leq 6)$.

Example 15. An automobile smog check is performed in three stages by a mechanic. Each stage is exponentially distributed with a mean service time of 7 minutes. Find the probability that

1. the first stage will take less than 5 minutes.
2. the first stage will take more than 12 minutes given that it has already lasted 7 minutes.
3. the entire exam will take 15 minutes or less.

Example 16. The time to pass through a queue to begin self-service at a buffet restraunt is represented by random variable T. What is the probability that an arriving customer will have to wait more than 15 minutes in the queue if T is modeled as

1. $N(18, 16)$?
2. $T(12, 18, 20)$?
3. $C(18, 16)$?

Conditional Expectation

Given random variables X and Y, the **conditional expectation of X given Y** is a real-valued random variable, denoted $E[X|Y]$. Moreover, the value of the random variable depends upon the value of Y. For example, if Y takes on the value $Y = y_j$, then

$$E[X|Y](y_j) = \sum_{i=1}^{m} x_i Pr(X = x_i | Y = y_j).$$

Law of Expectation of Conditional Expectation. $E[E[X|Y]] = E[X]$

Proof of Law of Expectation of Conditional Expectation.

$$E[E[X|Y]] = \sum_{j=1}^{n} (\sum_{i=1}^{n} x_i Pr(X = x_i | Y = y_j)) Pr(Y = y_j) =$$

$$\sum_{i=1}^{n} x_i (\sum_{j=1}^{n} Pr(X = x_i | Y = y_j) Pr(Y = y_j)) =$$

$$\sum_{i=1}^{n} x_i Pr(X = x_i) = E[X].$$

QED

Example 17: the buffon needle problem. A needle of length l is randomly dropped onto a plane that is divided by parallel lines, where each adjacent pair of lines are a distance l from one another. What is the Law of ECE to compute the probability that the needle will avoid crossing one of the lines.

Another remarkable property of conditional probability is that it allows for a joint density function to be factored into conditional densities. Indeed, if $X_1, X_2, \ldots, X_n$, $n \geq 2$, are random variables with joint density function $p(x_1, x_2, \ldots, x_n)$, then

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2) \cdots p(x_n|x_1, \ldots, x_{n-1}).$$

Counting Processes and a Relationship Between the Poisson and Exponential Distribution

Stochastic Process: a collection of random variables $\{X(t), t \in \mathcal{T}\}$. Usually $\mathcal{T}$ represents an interval on the real-line, such as $[0, T]$, for some $T > 0$.

Counting Process: a stochastic process $\{N(t), t \in \mathcal{T}\}$ for which $N(t)$ counts the "events" that have occurred up to time t.

Poisson Counting Process:

1. $N(0) = 0$
2. $N(t)$ has *independent increments*. In other words, the number of events occurring in $[s_1, t_1]$ is independent of the number of events occurring in $[s_2, t_2]$ when $[s_1, t_1] \cap [s_2, t_2] = \emptyset$
3. $N(t)$ has stationary increments. In other words, the distribution of the number of events occurring in $[s_1, t_1]$ is the same as the distribution for $[s_2, t_2]$ when $|t_1 - s_1| = |t_2 - s_2|$

Theorem For a Poisson counting process

1. for every $a \geq 0$, the distribution for the number of events occurring in interval $[a, a + t]$ is distributed as a Poisson random variable $P(\lambda t)$, for some $\lambda > 0$. λ is called the *arrival rate* of the process.
2. the elapsed time between two successive events has an exponential distribution with parameter λ.

Example 18. Suppose that people immigrate to a territory at a Poisson rate of $\lambda = 1$ per day. What is the expected time until the tenth immigrant arrives? What is the probability that the elapsed time between the tenth and eleventh arrival exceeds two days?

Exercises

1. Prove that if A and B are members of σ-algebra $\mathcal{E}$, then so is $A - B$.

2. Prove Statements 2 and 3 of Proposition 1.

3. Let P be a probability measure over σ-algebra $\mathcal{E}$, and suppose $E_1 \subseteq E_2 \subseteq \cdots$ is a nested sequence of sets, where $E_i \in \mathcal{E}$, for $i = 1, 2, \ldots$. Provide a summation formula for computing $P(\bigcup_{i=1}^{\infty} E_i)$.

4. Let $(S, \mathcal{E}, P)$ be a probability measure space. Show that, for measurable A and B satisfying $A \subseteq B$, $P(A) \leq P(B)$. This is called the **monotonicity property**.

5. Prove the Law of Total Probability.

6. Prove that if A and B are independent events, then so are $\overline{A}$ and B.

7. Let X be a random variable with domain $\mathcal{R}$, and suppose that $\mathcal{E}$ is a σ-algebra over X that contains all sets of the form $(-\infty, a]$, where a is a real number. In other words, it contains all events of the form $X \leq a$. Prove that $\mathcal{E}$ also contains the events $[a, b]$, $(a, b]$, $[a, b)$, and (a, b), for all real $a \leq b$.

8. Let X be a random variable with domain $\mathcal{R}$, and suppose that $\mathcal{E}$ is a σ-algebra over X that contains all sets of the form $(-\infty, a]$, where a is a real number. In other words, it contains all events of the form $X \leq a$. Use the previous exercise to prove that, if X is discrete, then $\mathcal{E}$ must equal the discrete algebra.

9. Prove that if X and Y are independent, then $E[XY] = E[X]E[Y]$. Hint: prove it for the case when X and Y are both finite.

10. For any constant a, prove that $\text{Var}(aX) = a^2\text{Var}(X)$.

11. Let S be the space of all permutations of the numbers $\{1, 2, 3, 4, 5\}$. Determine i) $|S|$. Also, let A denote the event that 1 occurs first in the permutation, B the event that 2 occurs second in the permutation, and C the event that 3 occurs third in the permutation. Determine the sizes of ii) $A \cap B$, iii) $A \cup B$, iv) $A \cap B \cap C$, and v) $A \cup (B \cap C)$.

12. Let $X = \{1, 2, 3, 4\}$, where $p(i) = ic$, for some constant c. Determine $P(2 \le X \le 3)$.

13. Determine $E[X]$ and $\text{Var}(X)$ for the random variable X defined in the previous exercise.

14. A machine at a children's arcade dispenses Harry Potter character stickers. There are 10 different character stickers. Assume that when a child inserts a token, the machine dispenses a single sticker that is randomly and uniformly chosen from the set of possible stickers. If a child successively inserts N tokens, how many different character stickers is she expected to have? Assume the N stickers are independently dispensed. Hint: let X_i, $i = 1, \ldots, 10$, be the indicator random variable that evaluates to 1 if the type-i character sticker is one of the N stickers, and evaluates to 0 otherwise. Then $X = \sum\limits_{i=1}^{10} X_i$.

15. Let $X = \{1, \ldots, 6\}$ have the uniform distribution. Compute $E[X]$ and $\text{Var}(X)$.

16. Let $X = [0, 1]$ with density function $f(x) = ce^x$. Compute $E[X]$ and $\text{Var}(X)$.

17. A bottle of commercial aloe vera water contains an average of 4 grams of aloe vera juice. i) What can be said about the probability that a randomly selected bottle contains 6 or more grams? ii) Same question, but, in addition, assume $\text{Var}(X) = 1(\text{grams})^2$.

18. Teams A and B are to play each other in a series of games. In any given game, Team A can beat Team B with probability $p > 0.5$. Prove that Team A has a better chance at winning a best-of-five series than it does at winning a best-of-three series. For example, in a best-of-three series, the first team to win two games wins the series.

19. For binomial random variable X distributed as $B(n, p)$, prove that $p_i = P(X = i)$ increases in i until it reaches its largest value when i is the greatest integer that is less than or equal to $(n + 1)p$. Hint: write a recurrence for p_{i+1} in terms of p_i.

20. If X is distributed as $B(m, p)$ and Y as $B(n, p)$, then how is $X + Y$ distributed. Justify your answer in a few sentences.

21. Suppose while typing an essay, a word is misspelled with probability equal to 0.002. Suppose the essay is to have 5,000 words. Let X count the number of spelling errors. If we assume X is distributed as $B(5,000, 0.002)$, then use a Poisson approximation to estimate $P(X \le 5)$.

22. Prove that if X is distributed as $\text{Poisson}(\lambda)$, then $E[X] = \lambda$.

23. If X and Y are independent Poisson random variables with respective parameters λ_1 and λ_2, then prove that $X + Y$ is Poisson with paramter $\lambda_1 + \lambda_2$.

24. If X is a Geometric random variable with parameter p, then compute $P(X > n)$, where $n > 0$ is an integer.

25. If two teams play a best-of-five series, determine the probability that Team A wins, assuming that the games have independent outcomes, and Team A wins a game with probability 0.6. Hint: use a negative binomial random variable.

26. Recall that $X = \{0, 1, \ldots, r\}$ has the hypergeometric distribution $\mathrm{HG}(m, n, r)$ when X counts the number of blue balls selected from an urn containing m blue balls, and n red balls, when a total of r balls are selected. Prove $E[X] = \frac{mr}{m+n}$. Hint: assume the blue balls are numbered 1 through m, and let $X_i = 1$ if blue ball i is selected, and zero otherwise.

27. A bus on a busy route is scheduled to reach a bus stop at 8:00AM every morning. Unfortunately, the bus is always T minutes late, where T has an exponential distribution with mean 3 minutes. If I arrive at $8:05$ AM, what is the probability that I missed this bus?

28. Repeat the previous exercise, but now assume that T is normally distributed with a mean of 3 minutes, and a variance of 2 minutes^2.

29. Prove that $E[X] = 1/\lambda$, where X is distributed as $E(\lambda)$.

30. One interesting property of an exponential random variable X is that it is **memoryless**, meaning that
$$P(X > s + t | X > s) = P(X > t).$$
For example, if X represents the number of seconds before the arrival of the next visitor to a website, then the probability that the next visit will occur more than t seconds later on condition that there have been no visits for the past s seconds, is equal to the unconditional probability that the next visit will be in t seconds. In other words, the likelihood of having to wait at least t more seconds for a visit is independent of the fact that we have already waited for s seconds without a visit. Prove that X is memoryless if it is exponentially distributed.

31. Customers A, B, and C enther a bank with two tellers. The tellers begin serving A and B, while C waits for the first teller to finish. If both tellers have service times that are exponentially distributed with parameter λ, explain why there is a 50% chance that C is the last customer to complete service. Hint: very little work is needed to answer this exercise.

32. Let X and Y be exponential random variables with respective parameters λ and μ. Is $\max(X, Y)$ also exponential?

33. The number of orders placed at a bakery over an hour period follows a Poisson random variable with rate $\lambda = 0.3$ orders/hour. Calculate the probability that no orders are placed between 10:00AM and 2:00PM.

34. Let $N(t)$ denote a Poisson counting process with rate λ events per unit time. For $t > s$ compute $P(N(t) = n | N(s) = m)$.

35. Repeat the previous problem, but now compute $P(N(s) = m | N(t) = n)$. In other words, given that n events have occurred over t units of time, what is the probability that m of them occurred during the first s units of time.

36. An urn contains 4 white balls and 6 black balls. Suppose 4 balls are randomly drawn from the urn. Let random variable X denote the number of white balls drawn. Suppose a 5th ball is drawn. Let $Y = 1$ if this ball is white, and 0 otherwise. Compute a) $E[Y | X = 2]$, b) $E[X | Y = 1]$, c) $\mathrm{Var}(Y | X = 0)$, d) $\mathrm{Var}(X | Y = 1)$.

37. Let X and Y be independent identically distributed exponential random variables. Prove that X is uniformly distributed over $[0, t]$ on condition that $X + Y = t$, for some constant $t > 0$.

Note: this result may seem somewhat counterintuitive. For example, if $\lambda = 1$ is the arrival rate for both X and Y, and $t = 2$, then one might assume that it is more likely for X to be in the interval $[1-\epsilon, 1+\epsilon]$ than in the interval $[0, 2\epsilon]$, since the former is closer to the mean of X and Y. But X is equally likely to occur in either one.

38. Prove that if $U \sim \mathcal{U}(0,1)$, then $\min(U, 1-U)$ is uniformly distributed over $[0, 1/2]$. Hint: assume that $U = \sum_{i=0}^{\infty} X_i 2^{-i}$, where $X_0, X_1, \ldots$ are independent and identically distributed Bernoulli random variables with $p = 1/2$. In other words, there is a one-to-one correspondence between numbers in $[0,1]$ and infinite binary sequences, and these numbers are uniformly distributed iff the bits of the binary sequence are generated by the independent tosses of a fair coin.

Exercise Solutions

1. Let S denote the universe containing $\mathcal{E}$. Then, since $\mathcal{E}$ is a σ-algebra, $S - B \in \mathcal{E}$. Therefore,

$$A - B = A \cap (S - B) \in \mathcal{E}.$$

2. From the previous exercise we know that $A - B$ is a member of the σ-algebra. Also, since $A = (A-B) \cup (A \cap B)$, and $A - B$ and $A \cap B$ are disjoint, then by Axiom 2, $P(A) = P(A-B) + P(A \cap B)$, which implies, $P(A-B) = P(A) - P(A \cap B)$. Similarly, $P(B-A) = P(B) - P(A \cap B)$. Finally, since $A - B$, $B - A$, and $A \cap B$ are all disjoint and union to $A \cup B$, by Axiom 2,

$$P(A \cup B) = P(A - B) + P(B - A) + P(A \cap B) =$$

$$(P(A) - P(A \cap B)) + (P(B) - P(A \cap B)) + P(A \cap B) = P(A) + P(B) - P(A \cap B).$$

To prove that $P(\overline{E}) = 1 - P(E)$, note that, since $S = E \cup \overline{E}$, $P(S) = P(E) + P(\overline{E}) = 1$, which implies $P(\overline{E}) = 1 - P(E)$.

3. Let $E_0 = \emptyset$, and $F_i = E_i - E_{i-1}$, $i = 1, 2, \ldots$. Then $F_1, F_2, \ldots$ are pairwise disjoint, with

$$\bigcup_{i=1}^{\infty} E_i = \bigcup_{i=1}^{\infty} F_i.$$

Thus,

$$P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(F_i).$$

4. Since $A \subseteq B$, we have $B = (B - A) \cup A$. And since A and $B - A$ are disjoint, it follows that

$$P(A) \leq P(A) + P(B - A) = P(B).$$

5. Suppose events $E_1, \ldots, E_n$ is a partition of X. Then $X = E_1 \cup \cdots \cup E_n$. Then

$$P(A) = P(A \cap X) = P(A \cap (E_1 \cup \cdots \cup E_n)) = P((A \cap E_1) \cup \cdots \cup (A \cap E_n)) =$$

$$\sum_{i=1}^{n} P(A \cap E_i) = \sum_{i=1}^{n} P(A|E_i)P(E_i),$$

where the third equality is from the distributive rule of $\cap$ over $\cup$, the fourth equality if from the fact that the E_i are disjoint, and the final equality is from the definition of conditional probability.

6. The result is immediate by definition if $P(B) = 0$. So assume $P(B) > 0$. Then

$$P(\overline{A} \cap B) = P(\overline{A}|B)P(B) = (1 - P(A|B))P(B) = (1 - P(A))P(B) = P(\overline{A})P(B),$$

where the second-to-last equality is from the assumption that A and B are independent.

7. Since $(-\infty, a]$ is in $\mathcal{E}$, it follows that the complement of $(-\infty, a]$, namely (a, ∞) is in $\mathcal{E}$, for all $a \in \mathcal{R}$. Moreover,

$$(a, b] = (a, \infty) \cap (-\infty, b]$$

is in $\mathcal{E}$, since $\mathcal{E}$ is closed under countable intersections. Furthermore,

$$[a, b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b\right],$$

and so $[a, b]$ is in $\mathcal{E}$. Finally,

$$(a, b) = \bigcup_{n=k}^{\infty} \left(a, b - \frac{1}{n}\right],$$

(where k is chosen large enough so that $b - 1/k \geq a$) is in $\mathcal{E}$, since $\mathcal{E}$ is closed under countable unions. A similar argument holds for sets of the form $[a, b)$.

8. It suffices to prove that every singleton set $\{a\}$ is an element of $\mathcal{E}$. This would imply that every discrete subset of the form $\{a_1, \ldots, a_n\}$ is in $\mathcal{E}$, since $\mathcal{E}$ is closed under countable unions. But from the previous exercise every closed interval $[a, b]$ is in $\mathcal{E}$. Therefore,

$$\{a\} = \bigcap_{n=1}^{\infty} \left[a - \frac{1}{n}, a + \frac{1}{n}\right]$$

is a member of $\mathcal{E}$, since $\mathcal{E}$ is closed under countable intersections.

9. Let $X = \{x_1, \ldots, x_m\}$ and $Y = \{y_1, \ldots, y_n\}$. Then

$$E[XY] = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j P(X = x_i, Y = y_j) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j P(X = x_i)P(Y = y_j) =$$

$$\sum_{i=1}^{m} x_i P(X = x_i) \sum_{j=1}^{n} y_j P(Y = y_i) = E[X]E[Y],$$

where the second equality is due to independence of X and Y, and the third is due to the fact that $x_i P(X = x_i)$ does not depend on the j summation index, and hence may be moved outside the second summation.

10. For any constant a,

$$\mathrm{Var}(aX) = E[(aX - E[aX])^2] = E[(aX - aE[X])^2] =$$

$$E[a^2(X - E[X])^2] = a^2 E[(X - E[X])^2] = a^2 \mathrm{Var}(X).$$

11. i) $|S| = 5! = 120$. ii) $|A \cap B| = 3! = 6$. iii)

$$|A \cup B| = |A| + |B| - |A \cap B| = 4! + 4! - 3! = 42.$$

iv) $|A \cap B \cap C| = 2! = 2$. v)

$$|A \cup (B \cap C)| = |A| + |B \cap C| - |A \cap B \cap C| = 24 + 6 - 2 = 28.$$

12. $P(2 \le X \le 3) = p(2) + p(3) = 2c + 3c = 5c$, where c must satisfy $10c = 1$. Therefore, $c = 0.1$, and $P(2 \le X \le 3) = 0.5$.

13. $E[X] = 1(0.1) + 2(0.2) + 3(0.3) + 4(0.4) = 3$ and

$$E[X^2] = 1^2(0.1) + 2^2(0.2) + 3^2(0.3) + 4^2(0.4) = 10.$$

Therefore, $\text{Var}(X) = E[X^2] - E^2[X] = 10 - 9 = 1$.

14. Since X_i is an indicator random variable, $E[X_i] = P(X_i = 1)$. But $P(X_i = 1)$ equals the probability that at least one of the N stickers is type i, which equals 1 minus the probability that none of the N stickers are type i. Since the N stickers are independently dispensed, the probability that none of them is type i equals $(0.9)^N$. Hence, $P(X_i = 1) = 1 - (0.9)^N$. Therefore,

$$E[X] = \sum_{i=1}^{10} E[X_i] = 10(1 - (0.9)^N).$$

15. $E[X] = \frac{1+2+\cdots+6}{6} = 7/2 = 3.5$.

$$E[X^2] = \frac{1^2 + 2^2 + \cdots + 6^2}{6} = (6)(7)(13)/(6 \cdot 6) = 91/6.$$

Therefore, $\text{Var}(X) = 91/6 - 49/4 = 35/12$.

16. Since $\int_0^1 ce^x = 1$, it follows that $c = 1/(e-1)$. Also,

$$E[X] = c\int_0^1 xe^x dx = c(xe^x|_0^1 - \int_0^1 e^x dx) = c,$$

while

$$E[X^2] = c\int_0^1 x^2e^x dx = c(x^2e^x|_0^1 - 2\int_0^1 xe^x dx) = c(e-2).$$

Therefore,

$$\text{Var}(X) = c(e-2) - c^2 = \frac{e^2 - 3e + 1}{(e-1)^2}.$$

17. Let X denote the number of grams of aloe vera juice measured in a randomly selected production bottle. i) By Markov's inequality,

$$P(X \ge 6) \le E[X]/6 = 4/6 = 2/3.$$

ii) Since $\text{Var}(X) = \sigma^2 = 1$, $\sigma = 1$, and Chebyshev's inequality yields

$$P(|X - 4| \ge (2)(1)) \le \frac{1}{2^2} = 1/4.$$

This is a much better bound than Markov's.

18. A best-of-three series can end with one of four possible events: $E_{2,0}$, $E_{2,1}$, $E_{1,2}$, and $E_{0,2}$. For example, $E_{2,0}$ is the event that Team A wins the first two games. Moreover, $P(E_{2,0}) = p^2$, $P(E_{2,1}) = 2p^2(1-p)$, $P(E_{0,2}) = (1-p)^2$,, and $P(E_{1,2}) = 2p(1-p)^2$. By symmetry, and since $p > 1-p$, we see that Team A has the better chance of winning a best-of-three series. Now we want to show that they have an even better chance of winning a best-of-five series.

    Now suppose that the teams agree to play best-of-three, but the losing team petitions the winning team to continue playing to a best-of-five. In other words, the best-of-three is extended to a best-of-five at the request of the losing team. We say that a team **reverses its outcome** iff it loses the best-of-three, but wins the best-of-five. We show that Team A has a better chance at reversing its outcome. To see this, first consider the event $E_{0,2}$ which has measure $(1-p)^2$. Then Team A can reverse a fraction p^3 of this measure, for a total measure of $p^3(1-p)^2$. Similarly, for the event $E_{1,2}$ which has measure $2p(1-p)^2$, team A can reverse a fraction p^2 of this measure for a total measure of $2p^3(1-p)^2$. Thus, by continuing to a best-of-five, Team A can reverse a total of $3p^3(1-p)^2$ of its best-of-three losing probability. Moreover, by symmetry, Team B can reverse a total of $3(1-p)^3p^2 < 3p^3(1-p)^2$ of its losing probability. This yields a net increase of winning probability for Team A. Therefore, Team A has an even better chance at winning a best-of-five series.

    Note: this argument can be generalized for any best-of-n series versus best-of-$n+2$ series. In general, a lesser opponent wants to play as few games as possible against a better opponent.

19. By definition, we have

$$p_{i+1} = \frac{n!}{(n-i-1)!(i+1)!}p^{i+1}(1-p)^{n-i-1} =$$

$$\frac{n!(n-i)}{(n-i)!(i+1)i!}p^i(1-p)^{n-i}\frac{p}{1-p} =$$

$$\frac{n-i}{i+1}\frac{p}{1-p}p_i.$$

    Thus $p_{i+1} \geq p_i$ iff

$$\frac{n-i}{i+1}\frac{p}{1-p} \geq 1 \Leftrightarrow$$

$$np - ip \geq (i+1)(1-p) \Leftrightarrow np - ip \geq -ip - p + 1 + i \Leftrightarrow$$

$$i + 1 \leq p(n+1).$$

20. By definition X counts the number of m independent experiments that result in success, where p is the probability of a single experiment resulting in success. Similarly, Y counts the number of n independent experiments that result in success. Therefore, $X + Y$ counts the number of $m + n$ experiments that result in success, where p is the probability of a single experiment resulting in success. In other words, $X + Y$ is distributed as $B(m+n, p)$.

21. We have $\lambda = np = (5000)(0.002) = 10$. Then

$$P(X \leq 5) = e^{-10}(1 + 10 + 100/2 + 1000/6 + 10000/24 + 100000/120) = 0.067.$$

    Therefore, there is over a 94% chance that the essay will have at least 6 spelling errors.

22. We have

$$E[X] = \sum_{n=0}^{\infty} n e^{-\lambda}\frac{\lambda^n}{n!} = e^{-\lambda}\lambda\sum_{n=1}^{\infty}\frac{\lambda^{n-1}}{(n-1)!} =$$

$$e^{-\lambda}\lambda\sum_{n=0}^{\infty}\frac{\lambda^n}{n!} = (e^{-\lambda}\lambda)(e^{\lambda}) = \lambda,$$

where the second-to-last equality uses the fact that $\sum_{n=0}^{\infty}\frac{x^n}{n!}$ is the Taylor series for e^x.

23. For integer $k \geq 0$.

$$P(X+Y=k) = \sum_{i=0}^{k} P(X=i, Y=k-i) = \sum_{i=0}^{k} P(X=i)P(Y=k-i) =$$

$$\sum_{i=0}^{k} e^{-\lambda_1}\frac{\lambda_1^i}{i!}e^{-\lambda_2}\frac{\lambda_2^{k-i}}{(k-i)!} = \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{i=0}^{k}\frac{k!}{(k-i)!i!}\lambda_1^i\lambda_2^{k-i} =$$

$$\frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{i=0}^{k}\binom{k}{i}\lambda_1^i\lambda_2^{k-i} = e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^k}{k!},$$

where the last equality uses the Binomial Theorem from discrete mathematics.

24. If X is a Geometric random variable with parameter p, then

$$P(X > n) = 1 - P(X \leq n) = 1 - \sum_{i=1}^{n}(1-p)^{i-1}p = 1 - \frac{p(1-(1-p)^n)}{1-(1-p)} =$$

$$1 - \frac{p(1-(1-p)^n)}{p} = 1 - (1-(1-p)^n) = (1-p)^n,$$

where the third equality uses the formula for summing a geometric series.

An alternative and much easier proof uses the meaning of X, namely that X represents the number of trials of an experiement until a success is observed. Then $P(X > n)$ is the probability that the first n experiments are all failures. Moreover, since $(1-p)$ is the probability of failure, and experiments are independent, we see that $P(X > n) = (1-p)^n$.

25. Let X count the number of games beyond three that must be played until Team A wins three games (assume the two teams continue playing until Team A wins three). Then the probability that Team A wins the best-of-five series is equal to $P(X=0)+P(X=1)+P(X=2)$, where X is distributed as $\text{NB}(3, 0.6)$. Thus,

$$P(X=0)+P(X=1)+P(X=2) = 0.6^3 + 3(0.4)(0.6^3) + 6(0.4^2)(0.6^3) = 0.68.$$

26. Assume the blue balls are numbered 1 through m, and let $X_i = 1$ if blue ball i is selected, and zero otherwise. Then

$$E[X] = E[\sum_{i=1}^{m} X_i] = \sum_{i=1}^{m} E[X_i].$$

But $E[X_i] = P(X_i = 1)$ is the probability that blue ball i is one of the r selected, which equals

$$\binom{m+n-1}{r-1} \Big/ \binom{m+n}{r} = \frac{r}{m+n}.$$

This is true, since $\binom{m+n-1}{r-1}$ is the number of r-combinations of balls that include blue ball i, while $\binom{m+n}{r}$ is the total number of r-combinations. Therefore, $E[X] = \frac{mr}{m+n}$.

27. The probability that I missed the bus is equal to

$$P(T < 5) = 1 - e^{-5/3} = 0.811.$$

So there's a good chance that I'll be catching the 8:30 bus.

28. The probability that I missed the bus is equal to

$$P(T < 5) = P(T - 3 < 2) = P(\frac{T-3}{\sqrt{2}} \leq 2/\sqrt{2}) =$$

$$P(Z < \sqrt{2}) = \Phi(\sqrt{2}) = 0.921,$$

where Z has the standard normal distirubtion.

29. We have

$$E[X] = \int_0^\infty \lambda x e^{-\lambda x} dx = x e^{-\lambda x}\Big|_0^\infty - \int_0^\infty e^{-\lambda x} dx =$$

$$0 - \frac{e^{-\lambda x}}{\lambda}\Big|_0^\infty = 1/\lambda.$$

30. If X is distributed as $E(\lambda)$, then $P(X > t) = e^{-\lambda t}$. Therefore,

$$P(X > s + t | X > s) = \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} =$$

$$\frac{e^{-\lambda s} e^{-\lambda t}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t).$$

31. Since A and B start at the same time, they are equally likely to finish first. So assume without loss of generality that A finishes first. Now, once C begins service at time t, B has already been in service for t units of time. But since the time to completion of B's service is memoryless (see previous exercise), it follows that, when C starts service, it is also as if B has also started service. Therefore, C is just as likely as B to finish last. Hence, C has a probability of 0.5 of finishing last.

32. Let $Z = \max(X, Y)$. Suppose $X < Y$, in that an X-arrival occurs at time t before a Y-arrival. Then

$$P(Z > s + t | Z > t) = P(Y > s + t | Y > t) = P(Y > s) = e^{-\mu s}.$$

Similarly, if $Y = t < X$ is observed at time t, then

$$P(Z > s + t | Z > t) = P(X > s + t | X > t) = P(X > s) = e^{-\lambda s}.$$

Thus, Z is exponentially distributed only if $\lambda = \mu$, and the arrival rate of Z is λ. But, assuming Z is the maximum of two exponential random variables with paramters λ, with probability one $Z > X$ which implies Z and X have different cumulative distributions. Therefore Z cannot be exponential with parameter λ. In general, $P(Z > s + t|Z > t) \neq P(Z > s)$ in cases where either X or Y has already been observed at or before time t.

33. We may count the number of orders placed between 10:00AM and 2:00PM using a single Poisson random variable X that has rate $\lambda = (4)(0.3) = 1.2$ (why?). Then $P(X = 0) = e^{-1.2} = 0.30$. Therefore, there is a 70% chance that someone will call in with an order between 0:00AM and 2:00PM.

34. We have $P(N(t) = n|N(s) = m) = 0$ if $n < m$, so assume $n \geq m$. Then we must compute the probability that an additional $n - m$ events will occur within a time period of $t - s$ time units. Moreover, since λ is the rate per unit time, it follows that $\lambda(t - s)$ is the rate for the $t - s$ time period. Thus,
$$P(N(t) = n|N(s) = m) = \frac{e^{-\lambda(t-s)}(\lambda(t-s))^{n-m}}{(n-m)!}.$$

35. Using the solution to the previous exercise, we have
$$P(N(s) = m|N(t) = n) = \frac{P(N(s) = m, N(t) = n)}{P(N(t) = n)} = \frac{P(N(s) = m)P(N(t) = n|N(s) = m)}{P(N(t) = n)} =$$
$$\left(\frac{e^{-\lambda s}(\lambda s)^m}{m!}\right)\left(\frac{e^{-\lambda(t-s)}(\lambda(t-s))^{n-m}}{(n-m)!}\right)\left(\frac{n!}{e^{-\lambda t}(\lambda t)^n}\right) =$$
$$\binom{n}{m}\left(\frac{s}{t}\right)^m\left(1 - \frac{s}{t}\right)^{n-m}.$$
The last expression admits an interesting interpretation. Given that n independent and random events $E_1, \ldots, E_n$ have occurred within t units of time, the probability that any one of these events occurs within s time units is equal to $\frac{s}{t}$. Thus, the probability that exactly m particular events $E_{i_1}, \ldots, E_{i_m}$ out of the n events have occurred is $(\frac{s}{t})^m(1 - \frac{s}{t})^{n-m}$. Finally, since there are $\binom{n}{m}$ different m-subsets of events, we arrive at the above formula. Thus, the rules of conditional probability support this more intuitive interpretation.

36. The solutions to each part are as follows.

   a. If $X = 2$, then this leaves 2 white balls and 4 black balls when the 5th ball is drawn. Thus,
   $$E[Y|X = 2] = P(Y = 1|X = 2) = 2/6 = 1/3.$$

   b. Now suppose $Y = 1$. Imagine removing a white ball and saving it later as the draw of the 5th ball. That leaves 3 white balls and 6 black balls for drawing the first four. Thus ,
   $$E[X|Y = 1] = \frac{(3)(4)}{3+6} = 4/3.$$

   c. If $X = 0$, then all 4 balls are black. This leaves 4 whites and 2 black for the 5th draw. And since Y is a Bernoulli random variable with $p = 2/3$, we have
   $$\text{Var}(Y|X = 0) = (2/3)(1/3) = 2/9.$$

d. Now suppose $Y = 1$. Imagine removing a white ball and saving it later as the draw of the 5th ball. That leaves 3 white balls and 6 black balls for drawing the first four. Thus, setting $m = 3$, $n = 6$, $r = 4$.

$$\mathrm{Var}(X|Y=1) = \frac{(3)(6)(4)(5)}{(9)^2(8)} = 5/9.$$

37. The joint density for (X, Y) is $f(x, y) = \lambda^2 e^{-\lambda x} e^{-\lambda y}$, $x, y \geq 0$. Now if we add the constraint $x + y = t$, or $y = t - x$, we get a density function $g(x)$ in terms of x equal to

$$g(x) = \frac{1}{c}\lambda^2 e^{-\lambda x} e^{-\lambda(t-x)} = \frac{1}{c}\lambda^2 e^{-\lambda t}$$

which is valid for $0 \leq x \leq t$, and where $c = \int_0^t \lambda^2 e^{-\lambda t} dx$. Notice that this density function does not depend on x. Therefore, X is uniformly distributed over $[0, t]$.

38. Using the hint that $U = \sum_{i=0}^{\infty} X_i 2^{-i}$, where $X_0, X_1, \ldots$ are independent and identically distributed Bernoulli random variables with $p = 1/2$, notice that

$$1 - U = \sum_{i=0}^{\infty}(1 - X_i)2^{-i} = \sum_{i=0}^{\infty} Y_i 2^{-i},$$

where $Y_0, Y_1, \ldots$ are independent and identically distributed Bernoulli random variables with $p = 1/2$. Thus, $\min(U, 1-U)$ chooses the sequence for which either $X_0 = 0$ or $Y_0 = 0$. WLOG, assume it chooses U. Then

$$\min(U, 1-U) = U = \sum_{i=1}^{\infty} X_i 2^{-i} = \frac{1}{2}\sum_{i=0}^{\infty} X_{i+1} 2^{-i}$$

is uniformly distirbuted over the interval $[0, 1/2]$.