Appendix B
Probability

Probability Spaces and Random Variables

A *probability space* is defined by a triple $(\Omega, \mathcal{F}, P)$, where Ω is a given set of elementary outcomes, $\mathcal{F}$ is a collection of subsets of Ω (each such subset B is called an *event*), and $P(\cdot)$ is a *probability measure* that assigns a nonnegative number $P(B)$ to each subset B in $\mathcal{F}$.

The collection of subsets $\mathcal{F}$ must satisfy

- If B is in $\mathcal{F}$, then so is its complement $\bar{B} = \{\omega \in \Omega : \omega \notin B\}$.
- If $B_1, B_2, \ldots$ are events in $\mathcal{F}$, then $\cup_k B_k$ and $\cap_k B_k$ are also in $\mathcal{F}$.

The probability measure must satisfy

- $P(B) \geq 0$ for all $B \in \mathcal{F}$.
- $P(\Omega) = 1$
- If $B_1, B_2, \ldots$ are disjoint events, the $P(\cup_k B_k) = \sum_k P(B_k)$.

A *random variable* is a function mapping elementary outcomes to real numbers, $X : \Omega \to \Re$ and is denoted $X(\omega)$—or simply X, where the dependence on ω is implicit. The *cumulative distribution function* (c.d.f.) of a random variable X—or just *distribution function* for short—is defined by

$$F(x) = P(X \leq x).$$

If X takes on only countable values, we define the *probability-mass function* (pmf) by the function

$$P(x) = P(X = x).$$

Such a random variable is said to be *discrete.* If F is differentiate, then the *probability-density function* is defined by

$$f(x) = \frac{\partial}{\partial x} F(x).$$

Such a random variable is said to be *continuous.*

Let $\mathbf{X} = (X_1, \ldots, X_n)$ denote a vector of random variables and $\mathbf{x} = (x_1, \ldots, x_n)$ a real vector. Then we define the *joint distribution* of X by

$$F(\mathbf{x}) = P(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

The random variables $X_1, \ldots, X_n$ are said to be *independent random variables* if

$$F(\mathbf{x}) = \prod_{i=1}^{n} F_i(x_i),$$

where $F_i(x_i) = P(X_i \leq x_i)$ (which is referred to as the *marginal distribution* of X_i). If, in addition, $F_i(x_i) = F(x_i)$ for all i, then the random variables are said to be *independent and identically distributed*—or *i.i.d.* for short.

Expectations and Moment-Generating Functions

The *expected value*—or *mean*—of a random variable X is defined by the integral

$$E[X] = \int_{\Re} x dF(x),$$

where the right-hand side above is equal to $\int_{\Re} x f(x) dx$ if X is continuous and $\sum_x x P\{X = x\}$ when X is discrete. For a general function $g : \Re \to \Re$, the expected value of $g(X)$ is defined by

$$E[g(X)] = \int_{\Re} g(x) dF(x).$$

The *variance* of X is defined as

$$Var(X) = E[(X - E[X])^2].$$

If X and Y are two random variables, the *covariance* is defined by

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])].$$

The *moment-generating function* of X is defined as

$$\psi_X(t) = E[e^{tX}].$$

The n^{th} *moment* of X is defined as $E[X^n]$. If the moment-generating function exists then one can determine the n^{th} moment of X using the fact that

$$E[X^n] = \frac{\partial^n}{\partial t} \psi_X(t)|_{t=0},$$

where $\frac{\partial^n}{\partial t}$ denotes the n^{th} derivative with respect to t.

The moment-generating function is also useful for analyzing sums of random variables. Indeed, if X and Y are two independent random variables with moment-generating functions $\psi_X(t)$ and $\psi_Y(t)$, respectively, and $Z = X + Y$, then

$$\psi_Z(t) = \psi_X(t)\, \psi_Y(t).$$

That is, the moment-generating function of a sum of independent random variables is simply the product of their individual moment-generating functions.

Inequalities

Jensen's inequality states that if g is a convex function, then

$$E[g(X)] \geq g(E[X]).$$

This is often useful in obtaining bounds on stochastic optimization problems.

Another useful bound in RM problems is due to Gallege [200] and involves a bound on the function $(X - x)^+ = \max\{X - x, 0\}$ (the positive part of $X - x$). It states that for any random variable X with mean μ and finite variance σ^2,

$$E[(X - x)^+] \leq \frac{\sqrt{\sigma^2 + (x - \mu)^2} - (x - \mu)}{2}.$$

For example, if X is demand and x is a capacity level, then $(X - x)^+$ is the rejected demand (spilled demand) and the above bound provides an upper bound on the expected spilled demand

Some Useful Distributions

We next provide the basic definitions of the most commonly used distributions in RM problems.

Discrete Distributions

Bernoulli

A random variable X has a Bernoulli distribution if it takes on only two values, 0 and 1. A Bernoulli distribution is characterized by a single parameter q (the probability that $X = 1$) with $0 \leq q \leq 1$. In RM, it is often used as the model of a single cancellation.

The basic definitions and properties are

$$\begin{aligned} P(x) &= \begin{cases} q & x = 1 \\ 1 - q & x = 0 \end{cases} \\ E[X] &= q \\ Var(X) &= q(1-q) \\ \psi(s) &= qe^s + (1-q). \end{aligned}$$

Binomial

A random variable X has a binomial distribution if it is the sum of n independent Bernoulli random variables. For example, the number of cancellations in a group of n reservations when each independently cancels with probability q. A binomial distribution is characterized by the two parameters q and n with $0 \leq q \leq 1$ and $n \geq 1$.

The basic definitions and properties are

$$\begin{aligned} P(x) &= \binom{n}{x} q^x (1-q)^{n-x}, \; x = 0, 1, \ldots, n \\ E[X] &= nq \end{aligned}$$

$$\begin{aligned} Var(X) &= nq(1-q) \\ \psi(s) &= (qe^s + (1-q))^n. \end{aligned}$$

Poisson

In RM, the Poisson distribution is used as a model of demand or as a (continuous parameter) approximation to the Binomial distribution. It is characterized by a single nonnegative parameter λ (its mean).

The basic definitions and properties are

$$\begin{aligned} P(x) &= e^{-\lambda}\frac{\lambda^x}{x!},\ x = 0, 1, \ldots \\ E[X] &= \lambda \\ Var(X) &= \lambda \\ \psi(s) &= e^{\lambda(e^s-1)}. \end{aligned}$$

Continuous Distributions

Uniform

A uniform distribution is defined by two constants $a < b$ and represents a case where the random variable is equally likely to assume any value in the interval $[a, b]$.

The basic definitions and properties are

$$\begin{aligned} f(x) &= \frac{1}{b-a},\ a \le x \le b \\ E[X] &= \frac{a+b}{2} \\ Var(X) &= \frac{(b-a)^2}{2} \\ \psi(s) &= \frac{e^{sb} - e^{sa}}{s(b-a)}. \end{aligned}$$

Exponential

The exponential distribution is defined by a single parameter λ.

The basic definitions and properties are

$$\begin{aligned} f(x) &= \lambda e^{-\lambda x},\ x \ge 0 \\ E[X] &= \frac{1}{\lambda} \\ Var(X) &= \frac{1}{\lambda^2} \\ \psi(s) &= \frac{\lambda}{\lambda - s}. \end{aligned}$$

Normal

The normal (or Gaussian) distribution is frequently used as a model of demand. It is characterized by two parameters, its mean μ and its variance σ^2.

The basic definitions and properties are

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty \\ E[X] &= \mu \\ Var(X) &= \sigma^2 \\ \psi(s) &= e^{\mu s + \frac{\sigma^2 s}{2}}. \end{aligned}$$

The normal has the property that if X and Y are two independent normal random variables, then the sum $X + Y$ also has a normal distribution (it is "closed under addition"). For example, if X and Y are independent with means μ_X, μ_Y and variances σ_X^2, σ_Y^2 (respectively), then their sum $X + Y$ has a normal distribution with mean $\mu_X + \mu_Y$ and variance $\sigma_X^2 + \sigma_Y^2$.

Gumbel

The Gumbel (or double-exponential) distribution is frequently used in discrete-choice models because it is "closed under maximization." That is, the maximum of two Gumbel random variables is also a Gumbel random variable. It is characterized by two parameters, a scale parameter μ and location parameter η.

The basic definitions and properties are

$$\begin{aligned} f(x) &= \frac{1}{\mu} e^{-\frac{x-\eta}{\mu}} e^{-e^{-\frac{x-\eta}{\mu}}} \quad -\infty < x < \infty \\ E[X] &= \eta + \frac{1}{\gamma\mu} \\ Var(X) &= \frac{\mu^2\pi^2}{6} \\ \psi(s) &= e^{\eta s/\mu}\Gamma(1 + s\mu), \end{aligned}$$

where $\gamma \approx= 0.577$ is Euler's constant and $\Gamma(x)$ is the extension of the factorial function to real numbers

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt.$$

If X_1 and X_2 are two independent Gumbel random variables with parameters (η_1, μ) and (η_2, μ) respectively, then $\max\{X_1, X_2\}$ is a Gumbel random variable with parameters $(\mu(\ln(e^{\eta_1/\mu} + e^{\eta_2/\mu}), \mu)$.

Stochastic Monotonicity and Convexity

Consider a random variable X that depends on some parameter θ, so that $X = X(\theta)$. That is, $X(\theta)$ is a *random function* of θ. For example, X could be the number of customers who show up out of θ reservations, in which case

$$X(\theta) = \sum_{i=1}^{\theta} Y_i,$$

where Y_i are i.i.d. Bernoulli random variables with $P(Y_i = 1) = q$ and $P(Y_i = 0) = 1 - q$.

Given a function $g(x)$, suppose we are interested in determining properties of the expected value $E[g(X(\theta)]$ as a function of θ. For example, if g is increasing in x, is $E[g(X(\theta)]$ increasing in θ? If g is convex in x, is $E[g(X(\theta)]$ convex in θ? Stochastic monotonicity and convexity identify classes of random variables $X(\theta)$ for which such statements can be made. A good source for this material is the series of papers by Shaked and Shantikumar [460, 461] and their subsequent book [462].

DEFINITION B.1 *The random function* $X(\theta)$ *is* **stochastically increasing** *in* θ *if for all* $\theta_1 \geq \theta_2$, $P(X(\theta_1) > x) \geq P(X(\theta_2) > x)$.

A random function $X(\theta)$ is *stochastically decreasing* in θ if—$X(\theta)$ is stochastically increasing. An equivalent definition is provided by the following proposition:

PROPOSITION B.1 $X(\theta)$ *is stochastically increasing in* θ *if for any* $\theta_1 \geq \theta_2$, *there exists two random variables* X_1 *and* X_2 *defined on a common probability space* $(\Omega, \mathcal{F}, P)$, *such that* X_1 *and* X_2 *are equal in distribution to* $X(\theta_1)$ *and* $X(\theta_2)$ *(respectively), and they satisfy* $X_1(\omega) \geq X_2(\omega)$ *for all* $\omega \in \Omega$.

Continuing our example, we see that if $X(\theta) = \sum_{i=1}^{\theta} Y_i$, where Y_i are i.i.d. Bernoulli random variables, then $X(\theta)$ is stochastically increasing, since we can consider ω to define an infinite sequence $\{Y_1, Y_2, \ldots\}$ and consider $X(\theta)$ to be the sum of the first θ variables in this sequence. For every $\theta_1 \geq \theta_2$, the sums $X(\theta_1)$ and $X(\theta_2)$ will have the required distribution and $X(\theta_1) \geq X(\theta_2)$ for every such sequence ω.

The following proposition follows easily from this sample path definition of monotonicity:

PROPOSITION B.2 $X(\theta)$ is *stochastically increasing in* θ *if and only if for any real valued, increasing function* $g(x)$, $E[g(X(\theta)]$ *is increasing in* θ.

Similarly, one can define a notion of stochastic convexity for $X(\theta)$:

DEFINITION B.2 $X(\theta)$ *is* **stochastically convex** *(SCX) if for any real valued, convex function* $g(x)$, $E[g(X(\theta)]$ *is convex in* θ.

We say $X(\theta)$ is *stochastically concave* (SCV) if $-X(\theta)$ is stochastically convex, and we say $X(\theta)$ is *stochastically linear* if it is both stochastically convex and stochastically concave.

To verify whether the above holds is often difficult. However, two stronger notions of stochastic convexity are quite useful and both imply stochastic convexity. These are:

DEFINITION B.3 $X(\theta)$ *is said to be* **strongly stochastically convex** *(SSCX) if* $X(\theta) = \psi(Z, \theta)$ *where Z is a random variable independent of* θ *and* ψ *is convex in* θ *for every value of Z.*

For example, suppose $X(\theta) = \sigma Z + \theta$, where Z is a standard normal random variable. Then $X(\theta)$ is normal with mean θ and standard deviation σ, and $X(\theta)$ is strongly stochastically convex in θ.

A somewhat weaker version of stochastic convexity is the following:

DEFINITION B.4 *$X(\theta)$ is* **stochastically convex in the sample-path sense** *(SCX-sp) if for any four values $\theta_i, i = 1,2,3,4$ satisfying $\theta_2 - \theta_1 = \theta_4 - \theta_3$ and $\theta_4 \geq \max\{\theta_2, \theta_3\}$, there exist random variable $X_i, i = 1,2,3,4$ defined on a common probability space $(\Omega, \mathcal{F}, P)$, such that X_i is equal in distribution to $X(\theta_i)$, $i = 1,2,3,4$ and*

$$X_4(\omega) - X_3(\omega) \geq X_2(\omega) - X_1(\omega),$$

for all $\omega \in \Omega$.

To illustrate, we show that the sum of Bernoulli random variables is stochastically convex (and concave) in this sample path sense. To do so, let $\theta_i, i = 1,2,3,4$ be integers satisfying $\theta_2 - \theta_1 = \theta_4 - \theta_3$ and $\theta_4 \geq \max\{\theta_2, \theta_3\}$, and let ω define an infinite sequence $\{Y_1, Y_2, \ldots\}$ of i.i.d. Bernoulli random variables as before. Note that $\theta_1 \leq \min\{\theta_2, \theta_3\}$ (else $\theta_4 < \max\{\theta_2, \theta_3\}$), and define

$$\begin{aligned} X_1 &= \sum_{i=1}^{\theta_1} Y_i \\ X_3 &= \sum_{i=1}^{\theta_3} Y_i \\ X_4 &= \sum_{i=1}^{\theta_4} Y_i \\ X_2 &= X_1 + (X_4 - X_3). \end{aligned}$$

Note X_i is equal in distribution to $X(\theta_i)$ since each is the sum of θ_i i.i.d. Bernoulli random variables, and by construction

$$X_4 - X_3 = X_2 - X_1,$$

so $X(\theta)$ is stochastically convex in the sample path sense.

The following proposition relates these versions of stochastic convexity:

PROPOSITION B.3 *$SSCX \Rightarrow SICX\text{-}sp \Rightarrow SCX$.*

So showing $X(\theta)$ is either strongly stochastically convex or stochastically convex in the sample path sense, implies that $X(\theta)$ is stochastically convex. Again, returning to our example, this implies that if $X(\theta)$ is the sum of θ i.i.d. Bernoulli random variables and $g(x)$ is a convex function, the $E[g(X(\theta))]$ is convex in θ.