Stat 620: Applied Probability and Stochastic Models

1. Review of Probability

- An event E is a subset of a sample space $\mathbb{S}$.
- The probability of E is written as $\mathbb{P}(E)$.
- Formally, $\mathbb{P}(E)$ may be undefined for certain **non-measurable** events. Here, we assume $\mathbb{P}(E)$ is defined for all events of interest.

Axioms of Probability

A1. $\boxed{0 \leq \mathbb{P}(E) \leq 1}$

A2. $\boxed{\mathbb{P}(\mathbb{S}) = 1}$

A3. If $E_1, E_2, \ldots$ are mutually exclusive, then

$$\boxed{\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)}$$

Recall that $E_1, E_2, \ldots$ are **mutually exclusive** or **disjoint** if $E_i E_j = \emptyset$ for $i \neq j$, where $\emptyset$ is the **null event** or **empty set**.

Basic Consequences of the Axioms

B1. $\boxed{\mathbb{P}(\emptyset) = 0}$ Proof:

B2. For $E_1, \ldots, E_n$ mutually exclusive,
$\boxed{\mathbb{P}\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n \mathbb{P}(E_i)}$ Proof:

B3. If $E \subset F$ then $\mathbb{P}(E) \leq \mathbb{P}(F)$. Proof:

B4. $\boxed{\mathbb{P}(E^c) = 1 - \mathbb{P}(E)}$ where E^c is the **complement** of E. Proof:

B5. $\boxed{\mathbb{P}\left(\bigcup_{i=1}^\infty E_i\right) \leq \sum_{i=1}^\infty \mathbb{P}(E_i)}$. This is **Boole's inequality**. The proof is an exercise.

Continuity of Probability

- Let $\{F_n, n \geq 1\}$ be an **increasing** sequence of events, i.e., $F_n \subset F_{n+1}$ for all n.
- Define $\lim_{n\to\infty} F_n = \bigcup_{n=1}^{\infty} F_n$ for all n.
- Writing $E_1 = F_1$ and $E_n = F_n F_{n-1}^c$ for $n > 1$, A3 can be re-written as

B6. $\boxed{\mathbb{P}(\lim_{n\to\infty} F_n) = \lim_{n\to\infty} \mathbb{P}(F_n)}$

- $\{F_n\}$ is **decreasing** if $F_n \supset F_{n+1}$ for all n.
- In this case, $\lim_{n\to\infty} F_n = \bigcap_{n=1}^{\infty} F_n$. Since $\{F_n^c\}$ is increasing, and $\lim_{n\to\infty} F_n^c = (\lim_{n\to\infty} F_n)^c$ (why?), B6 also applies to decreasing sequences.

Why are we interested in $\lim_{n\to\infty}$?

- For small stochastic systems, one can calculate probabilities exactly.
- Large systems sometimes display simple limiting behavior despite the complexities of the full system.

Example: Statistical mechanics

- Solving laws of motion for as few as three interacting particles can be fiendishly difficult.
- The combined properties of 10^{23} particles are relatively tractable.

Example (Borel-Cantelli Lemma)

For any sequence $\{E_n, n \geq 1\}$ of events, if $\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty$ then $\mathbb{P}\big(\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} E_i\big) = 0$.

Note: $\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} E_i$ is the event that an infinite number of the E_i occur. Why?

Proof

Independence

- E_1 and E_2 are **independent** if $\mathbb{P}(E_1 E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2)$.
- $\{E_i, i = 1, \ldots, n\}$ are **independent** if any sequence $1 \leq i_1 < \ldots < i_k \leq n$ has $\mathbb{P}\left(\bigcap_{j=1}^k E_{i_j}\right) = \prod_{j=1}^k \mathbb{P}(E_{i_j})$.
- $\{E_i, i \geq 1\}$ are **independent** if any finite collection are independent.

Example (partial converse to Borel-Cantelli)

If $\{E_i, i \geq 1\}$ are independent, and $\sum_{i=1}^\infty \mathbb{P}(E_i) = \infty$, then

$\mathbb{P}(\text{infinite number of } E_i \text{ occur}) = 1$.

Proof

Proof continued

Random Variables

- A **random variable** X is a function from $\mathbb{S}$ to the real numbers, $X : \mathbb{S} \to \mathbb{R}$.
- $X \in A$ denotes the event $\{s : X(s) \in A\}$.
- The **distribution function** (or **cumulative distribution function** or **c.d.f.**) of X is $F_X(x) = \mathbb{P}(X \le x)$, so $F_X : \mathbb{R} \to [0, 1]$.
- RVs get capital letters, e.g. X, Y, Z, and their values get lower case, e.g. x, y, z.
- We write $\overline{F}_X(x) = 1 - F_X(x) = \mathbb{P}(X > x)$.
- For **jointly defined** RVs, e.g. X and Y, $F_{XY}(x, y) = \mathbb{P}(X \le x, Y \le y)$. Then, $F_X(x) = \lim_{y \to \infty} F_{XY}(x, y)$. Why?

Continuous Random Variables

- X is **continuous** if there is a **probability density function** (**p.d.f.**) $f_X(x)$ such that $\mathbb{P}(X \in A) = \int_A f_X(x)\,dx$.
- Since $F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx$, we have $f_X(x) = \frac{d}{dx} F_X(x)$. Why?

- X and Y are **jointly continuous** if they have a **joint p.d.f.** $f_{XY}(x, y)$ such that $\mathbb{P}(X \in A, Y \in B) = \int_A \int_B f_{XY}(x, y)\,dy\,dx$.

Discrete Random Variables

- X is **discrete** if it has a finite or countable set of **possible values** $x_1, x_2, \ldots$ in which case the distribution of X is determined by the **p.m.f.** (**probability mass function**) $p_X(x) = \mathbb{P}(X{=}x)$.
- Similarly, $p_{XY}(x, y) = \mathbb{P}(X{=}x, Y{=}y)$.

Independence of Random Variables

- Joint RVs $X_1, \ldots, X_n$ are **independent** if the events $\{a_1 \le X_1 \le b_1\}, \ldots, \{a_n \le X_n \le b_n\}$ are independent for all choices $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$.

Equivalently

$$F_{X_1 \ldots X_n}(x_1, \ldots, x_n) = \prod_{k=1}^{n} F_{X_k}(x_k)$$

Equivalently (For Continuous Random Variables)

$$f_{X_1 \ldots X_n}(x_1, \ldots, x_n) = \prod_{k=1}^{n} f_{X_k}(x_k)$$

Equivalently (For Discrete Random Variables)

$$p_{X_1 \ldots X_n}(x_1, \ldots, x_n) = \prod_{k=1}^{n} p_{X_k}(x_k)$$

- Checking these equivalences is an exercise!

- $X_1, \ldots, X_n$ are **independent and identically distributed** (**iid**) copies of X if they are independent and have the same **marginal** distribution as X, meaning $F_{X_i}(x) = F_X(x)$.

Expectation

- The **expected value** or **mean** of X is

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} x\, dF_X(x) && \text{[definition]} \\ &= \int_{-\infty}^{\infty} x f_X(x)\, dx && \text{[continuous RVs]} \\ &= \sum_x x p_X(x) && \text{[discrete RVs]} \end{aligned}$$

- $\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(x)\, dF_X(x)$

- Expectation is **linear**, i.e.

  $\mathbb{E}[\sum_{i=1}^{n} \lambda_i X_i] = \sum_{i=1}^{n} \lambda_i \mathbb{E}[X_i]$

- If X and Y are independent,

  $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$.

  The converse is not generally true.

Example: indicator random variables

For an event A, the **indicator random variable** I_A is defined as

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{else} \end{cases}$$

Then, $\mathbb{E}[I_A] = 1 \times \mathbb{P}(A) + 0 \times \mathbb{P}(A^c) = \mathbb{P}(A)$.

Variance and Covariance

- $\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$.
- $\mathrm{Cov}(X,Y) = \mathbb{E}\left[\left(X - \mathbb{E}[X]\right)\left(Y - \mathbb{E}[Y]\right)\right]$.
- Two useful identities:

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[X^2] - \left(\mathbb{E}[X]\right)^2 \\ \mathrm{Cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]\end{aligned}$$

  To check, expand & use linearity (exercise).

- If $\mathrm{Cov}(X,Y) = 0$, X and Y are **uncorrelated** (independence $\Rightarrow$ uncorrelated)

- Covariance is **bilinear**:

$$\mathrm{Cov}\left(\sum_{i=1}^m \lambda_i X_i, \sum_{j=1}^n \mu_j Y_j\right) = \sum_{i=1}^m \sum_{j=1}^n \lambda_i \mu_j \mathrm{Cov}(X_i, Y_j)$$

- Noting $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$ and $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$, this implies

$$\mathrm{Var}\left(\sum_{i=1}^m X_i\right) = \sum_{i=1}^m \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)$$

Integral Transforms

- If X has c.d.f. $F(x)$ then define

$\psi(t) = \mathbb{E}[e^{tX}] = \int e^{tx}\, dF(x)$ moment generating function (MGF)

$\phi(t) = \mathbb{E}[e^{itX}] = \int e^{itx}\, dF(x)$ characteristic function

$\widetilde{F}(t) = \mathbb{E}[e^{-tX}] = \int e^{-tx}\, dF(x)$ Laplace transform

- **uniqueness**: all three transforms uniquely determine $F(x)$ when they exist (we use this result without proof).

- **transforms of sums of independent RVs**: Let X and Y be independent. Then,

$$\begin{aligned}\psi_{X+Y}(t) &= \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX}]\,\mathbb{E}[e^{tY}] \\ &= \psi_X(t)\,\psi_Y(t)\end{aligned}$$

Where is independence used? Similarly,
$\phi_{X+Y}(t) = \phi_X(t)\,\phi_Y(t)$, $\widetilde{F}_{X+Y}(t) = \widetilde{F}_X(t)\,\widetilde{F}_Y(t)$.

- The MGF is useful for finding moments. The characteristic function always exists. Laplace transforms are convenient for non-negative RVs.

Some Standard Discrete Distributions

- $X \sim$ **Binomial** (n, p) if X counts the # of successes in n independent trials each with chance p of success.
- $X \sim$ **Poisson** (λ) if X counts the # of occurrences of rare, independent events (e.g. radioactive decay measured by a Geiger counter). This is a limit of Binomial (n, p) as $n \to \infty$ and $np \to \lambda$.
- $X \sim$ **Geometric** (p) if X counts the # of trials until the first success, in an infinite sequence of independent trials each with chance p of success.
- X is negative binomial, **Neg Bin** (r, p), if X counts the # of trials until r successes have occurred.
- The p.m.f., mean, variance and moment generating functions are on p16 of Ross. They may be used without proof in this course.

Example: Find the MGF of the Binomial (n, p) distribution, and hence find its mean & variance.

Some Standard Continuous Distributions

- $X \sim$ **Uniform** $[a, b]$, also written $X \sim \mathrm{U}[a, b]$, if

$$f_X(x) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & \text{else} \end{cases}.$$

- $X \sim$ **Exponential** (λ) if $f_X(x) = \lambda e^{-\lambda x}$ for $x \ge 0$ and $f_X(x) = 0$ for $x < 0$.

  The exponential is the only continuous distribution with the **memoryless property**: $\mathbb{P}(X > a + b \mid X > a) = \mathbb{P}(X > b)$ for $a > 0$, $b > 0$.

- $X \sim$ **Gamma** (n, λ) if $f_X(x) = \lambda^n x^{n-1} e^{-\lambda x}/(n-1)!$ for $x \ge 0$.

  If n is an integer, the Gamma distribution corresponds to the sum of n iid Exponential (λ) random variables.

- $X \sim$ **Normal** (μ, σ^2), or $X \sim \mathrm{N}(\mu, \sigma^2)$, if $f_X(x) = \frac{1}{2\pi\sigma^2} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}$.

- Z is **standard normal** if $Z \sim N(0, 1)$. We can write $X = \mu + \sigma Z$.

- $X \sim \textbf{Beta}\,(a,b)$ if $f_X(x) = cx^{a-1}(1-x)^{b-1}$ for $0 < x < 1$ where $c = \Gamma(a+b)/\Gamma(a)\Gamma(b)$ and $\Gamma(a)$ is the **gamma function** (if a is an integer, $\Gamma(a) = (a-1)!$).

Example If $U_1, \ldots, U_n \sim \text{iid U}\,[0,1]$, define $U_{(1)} < U_{(2)} < \ldots < U_{(n)}$ to be the **order statistics**, placing $U_1, \ldots, U_n$ in increasing order. Show that $U_{(k)} \sim \text{Beta}\,(k, n-k+1)$.

Solution

Solution continued

Conditional Probability and Expectation

- For events E and F, $\boxed{\mathbb{P}[E \,|\, F] = \mathbb{P}[EF] \,/\, \mathbb{P}(F)}$
- The **conditional c.d.f.** of X given $Y{=}y$ is $F_{X|Y}(x \,|\, y) = \mathbb{P}[X \leq x \,|\, Y{=}y]$ if $\mathbb{P}(Y{=}y) > 0$.
- The **conditional expectation** of X given $Y{=}y$ is $\mathbb{E}[X \,|\, Y{=}y] = \int x \, dF_{X|Y}(x \,|\, y)$.
- If X and Y are jointly continuous, the **conditional density** of X given $Y{=}y$ is $f_{X|Y}(x \,|\, y) = \frac{f_{XY}(x,y)}{f_Y(y)}$. Then, $F_{X|Y}(x \,|\, y) = \int_{-\infty}^{x} f_{X|Y}(z \,|\, y) \, dz$.
- Although we are conditioning on an event of probability zero, conditional densities behave like conditional probabilities
- $\mathbb{E}[X \,|\, Y{=}y] = \int x f_{X|Y}(x \,|\, y) \, dx$

Conditional Expectation as a Random Variable

- $\mathbb{E}[X \mid Y]$ is the RV which takes the value $\mathbb{E}[X \mid Y=y]$ when $Y=y$.
- $\boxed{\mathbb{E}\big(\mathbb{E}[X \mid Y]\big) = E[X]}$

  Proof: (For X and Y jointly continuous)

Other properties with similar proofs:

- $\boxed{\mathbb{E}[X\,h(Y) \mid Y] = h(Y)\,\mathbb{E}[X \mid Y] \text{ for any } h(\cdot)}$
- $\boxed{\mathbb{E}\big(\mathbb{E}[X \mid Y, Z] \mid Y\big) = \mathbb{E}[X \mid Y]}$

Conditional Probability of an Event Given a Random Variable

- All properties of conditional expectation apply also to conditional probabilities, since (by definition)

  $$\mathbb{P}[A \mid Y = y] = \mathbb{E}[I_A \mid Y = y]$$

  so we define

  $$\mathbb{P}[A \mid Y] = \mathbb{E}[I_A \mid Y]$$

  and we get

  $$\boxed{\mathbb{P}[A] = \mathbb{E}\big(\mathbb{P}[A \mid Y]\big)}$$

  Why?

- Note that this can also be written as

  $$\boxed{\mathbb{P}[A] = \int \mathbb{P}[A \mid Y = y] \, dF_Y(y)}$$

Example: Show that $h(Y)$ and $X - \mathbb{E}[X \mid Y]$ are uncorrelated, for an function $h(\cdot)$.

Solution

Example (Problem Solving Via Conditioning)

A gambler wins or loses $\$1$ with equal probability. She starts with $\$i$ and plays repeatedly until reading k or going broke. Find the expected number of times she plays.

Solution

Solution continued

Example: (least square prediction). If Y is used to predict X, we may choose the predictor $h(Y)$ to minimize $\mathbb{E}\big[(X - h(Y))^2\big]$. Show that this **least squares predictor** is $h(Y) = \mathbb{E}[X \mid Y]$.

Solution

Two Limit Theorems (without proofs)

Strong Law of Large Numbers: If $X_1, X_2, \ldots$ are independent & identically distributed (iid) with $\mathbb{E}[X_1] = \mu$, then

$$\mathbb{P}\left[\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu\right] = 1.$$

Central Limit Theorem: If $X_1, X_2, \ldots$ are iid with $\mathbb{E}[X_1] = \mu$ and $\mathrm{Var}[X_1] = \sigma^2$ then

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \leq a\right] = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

- Note that $\int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \mathbb{P}[Z \leq a]$ where $Z \sim N[0,1]$.

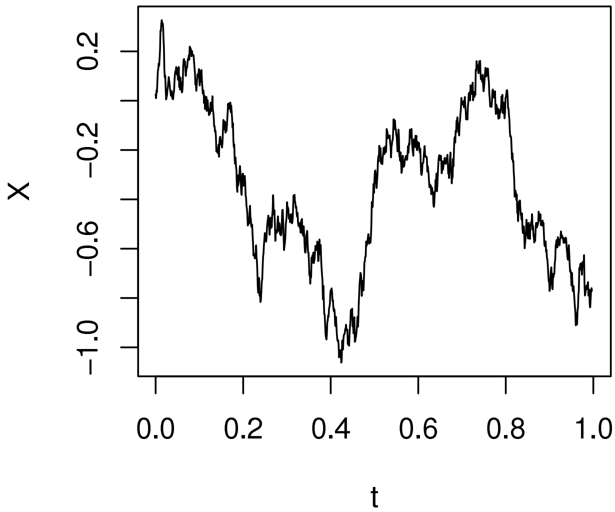
- The CLT can be checked experimentally (e.g., for a finite n one can simulate on a computer to see how closely the rescaled average matches the Normal distribution).

Stochastic Processes

- A **stochastic process** is a collection of random variables $\{X(t)\} = \{X(t), t \in T\}$ with an **index set** T.
- If T is countable, $\{X(t)\}$ is a **discrete-time** process. If T is an interval of $\mathbb{R}$, then $\{X(t)\}$ is a **continuous-time** process.
- The sample space for $\{X(t)\}$ is typically the set of all possible trajectories $\{x(t), t \in T\}$, termed **sample paths**.
- A continuous-time process may have continuous or discontinuous sample paths.
- $\{X(t)\}$ is **stationary** if $X(t_1), X(t_2), \ldots, X(t_k)$ has the same distribution as $X(t_1 + s), \ldots, X(t_k + s)$ for all choices of $t_1, \ldots, t_k$ and s. This means joint distributions are invariant to time shifts.

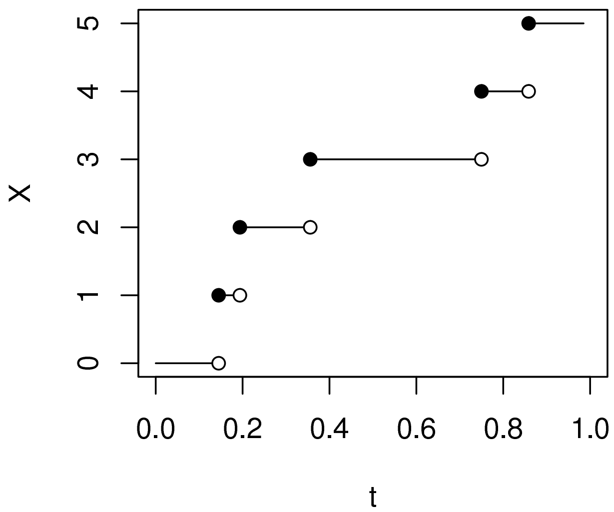
Simulating stochastic processes

- A process with continuous sample paths:




```
np=1000
t=seq(from=0,to=1,length=np)
X=cumsum(rnorm(np,mean=0,
  sd=sqrt(1/np)))
plot(t,X,type="l")
```

- Discontinuous sample paths: a jump process



Note: R code is not formally part of this course.

```
lambda=5
np=lambda
S=cumsum(rexp(np,rate=lambda))
plot(S,1:np,xlim=c(0,1),
  xlab="t",ylab="X",
  ylim=c(0,np),pch=19)
points(S,0:(np-1),pch=1)
matlines(rbind(
  c(0,S),c(S,1)-0.015),
  rbind(0:np,0:np),
  lty="solid",
  col="black")
```

- $\{X(t)\}$ has **stationary increments** if the distribution of $X(t+s) - X(t)$ doesn't depend on t.

- $\{X(t)\}$, has **independent increments** if $X(t_2)-X(t_1)$, $X(t_3)-X(t_2)$, ... , $X(t_k)-X(t_{k-1})$ are independent for all $t_1 < t_2 < \ldots < t_k$.

- $\{N(t)\}$ is a **counting process** if:
(i) $N(t)$ takes values in $\{0, 1, 2, 3, \ldots\}$, the non-negative integers.
(ii) $N(t)$ is increasing, i.e. for $s < t$, $N(s) \leq N(t)$.

- $N(t) - N(s)$ can be interpreted as counting the number of events arising between times s and t.

- Remarkably, there is essentially only one counting process with independent, stationary increments—the Poisson process. This process is widely used to model phenomena in natural sciences, engineering and social sciences. It is also a tool for building more complex models.