Final Review

MATH/STAT 394: Probability I
Summer 2021 A Term

Introduction to Probability
D. Anderson, T.Seppäläinen, B. Valkó

§ 3, 4

Aaron Osgood-Zimmerman

Department of Statistics

Logistics

Course evaluation

- Course evaluation avail. at `https://uw.iasystem.org/survey/238293`
- Please make any constructive comments for future courses

Outline

Random variables

Random variable A r.v. is a function from a prob. space $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathbb{R}$[1]

Discrete random variable

- A r.v X is **discrete** if it takes values in a finite/countable set $\mathcal{X}$ s.t. $\sum_{k \in \mathcal{X}} \mathbb{P}(X = k) = 1$
- A discrete r.v. is entirely characterized by its **p.m.f.**

$$p(k) = \mathbb{P}(X = k) \quad \text{for any } k \in \mathcal{X}$$

Continuous random variable

- A r.v. is **continuous** if there exists a function f, called the **p.d.f.** of X, s.t.

$$\mathbb{P}(X \in [a, b]) = \int_a^b f(x)dx \quad \text{for any } -\infty \leq a \leq b \leq +\infty$$

- A function f is a **valid p.d.f.**[2] if

$$\forall x \in \mathbb{R} \quad f(x) \geq 0, \qquad \int_{-\infty}^{+\infty} f(x)dx = 1$$

[1] Precisely, the pre-image of any Borel set must belong to $\mathcal{F}$

[2] Recall that changing the p.d.f. on a finite set of points does not change the computation of $\mathbb{P}(X \in [a, b])$, hence one r.v. can have multiple p.d.f. (up to a finite set of points)

Cumulative Distribution Function

c.d.f. defined for any r.v. X as

$$F(t) = \mathbb{P}(X \leq t)$$

From p.m.f./p.d.f. to c.d.f.

$$F(t) = \sum_{\substack{k \in \mathcal{X} \\ k \leq t}} p(k) \qquad \text{(for discrete r.v.)}$$

$$F(t) = \int_{-\infty}^{t} f(x)dx \qquad \text{(for continuous r.v.)}$$

From c.d.f. to p.m.f./p.d.f.

$$p(k) = F(k+) - F(k-) \qquad \text{(for discrete r.v.)}$$

$$f(x) = F'(x) \qquad \text{(for continuous r.v.)}$$

Cumulative Distribution Function

Charac. of random variables by the c.d.f.

- $F(t) = \mathbb{P}(X \leq t)$ is
  - non-decreasing
  - $F(-\infty) = 0$, $F(+\infty) = 1$
  - right continuous
  - $F(t-) = \mathbb{P}(X < t)$
- If F is piece-wise constant
  $\rightarrow$ it is the c.d.f. of a **discrete r.v.**
- If F is continuous
  $\rightarrow$ it is the c.d.f. of a **continuous r.v.**
- If F is discontinuous and not piece-wise constant
  $\rightarrow$ neither discrete nor continuous r.v.
  Yet, we can compute prob. using the c.d.f.

Expectation

Expectation

- **Definition**

$$\mathbb{E}[X] = \sum_{k \in \mathcal{X}} kp(k) \qquad \text{(for discrete r.v.)}$$

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} xf(x)dx \qquad \text{(for continuous r.v.)}$$

- **Expectation of a function of r.v.**

$$\mathbb{E}[g(X)] = \sum_{k \in \mathcal{X}} g(k)p(k) \qquad \text{(for discrete r.v.)}$$

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx \qquad \text{(for continuous r.v.)}$$

- An expectation can be finite, infinite (positive or negative) or undefined
- For 1_A the indicator r.v. of an event A,

$$\mathbb{E}[1_A] = \mathbb{P}(A)$$

- **Linearity of the expectation**

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b, \qquad \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

for any r.v. X, Y and $a, b \in \mathbb{R}$

Median Quantiles

Median Quantiles

- The **median** of a r.v. X is a value m s.t.

$$\mathbb{P}(X \geq m) \geq 1/2 \qquad \mathbb{P}(X \leq m) \geq 1/2$$

Note that for a continuous r.v., the median is unique and defined by

$$\mathbb{P}(X \geq m) = \mathbb{P}(X \leq m) = 1/2$$

- The **pth quantile** of a r.v. X is a value x_p s.t.

$$\mathbb{P}(X \leq x_p) \geq p \qquad \mathbb{P}(X \geq x_p) \geq 1 - p$$

Note that for a continuous r.v., the pth quantile is unique and defined by

$$\mathbb{P}(X \leq x_p) = p \qquad \mathbb{P}(X \geq x_p) = 1 - p$$

Variance

Variance

- **Variance**
$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}[X])^2\right) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$
- **Standard deviation** $\sqrt{\mathrm{Var}(X)}$
- For $a, b \in \mathbb{R}$
$$\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X)$$
- $\mathrm{Var}(X) = 0 \iff \mathbb{P}(X = a) = 1$ for some $a \in \mathbb{R}$

Classical discrete random variables

Classical r.v.

- $X \sim \mathrm{Ber}(p)$ for $p \in [0,1]$, if

$$\mathbb{P}(X=1)=p \quad P(X=0)=1-p$$

$$\text{so} \quad \mathbb{E}[X]=p, \quad \mathrm{Var}(X)=p(1-p)$$

- $X \sim \mathrm{Bin}(n,p)$ for $n \in \mathbb{N}\setminus\{0\}$, $p \in [0,1]$, if

$$\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, \quad \text{for } k=0,1,\dots,n$$

$$\text{so} \quad \mathbb{E}[X]=np, \quad \mathrm{Var}(X)=np(1-p)$$

- $X \sim \mathrm{Geom}(p)$ for $p \in [0,1]$, if

$$\mathbb{P}(X=k)=(1-p)^{k-1}p \quad \text{for } k=0,1,\dots$$

$$\text{so} \quad \mathbb{E}[X]=1/p, \quad \mathrm{Var}(X)=(1-p)/p^2$$

Classical discrete r.v.

- $X \sim \text{Poisson}(\lambda)$ for $\lambda > 0$ if

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for } k = 0, 1, \ldots$$

$$\text{so} \quad \mathbb{E}[X] = \lambda, \quad \text{Var}(X) = \lambda$$

- $X \sim \text{Hypergeom}(N, N_A, n)$,

$$\mathbb{P}(X = k) = \frac{\binom{N_A}{k}\binom{N-N_A}{n-k}}{\binom{N}{n}},$$

Classical continuous r.v.

- $X \sim \text{Unif}([a,b])$ for $a < b$, if it has a p.d.f.

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{if } x \notin [a,b] \end{cases}$$

$$\text{so} \quad \mathbb{E}[X] = \frac{a+b}{2} \quad \text{Var}(X) = \frac{(b-a)^2}{12}$$

- $X \sim \text{Exp}(\lambda)$ for $\lambda > 0$, if it has a p.d.f.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

$$\text{so} \quad \mathbb{E}[X] = 1/\lambda \quad \text{Var}(X) = 1/\lambda^2$$

- $X \sim \mathcal{N}(\mu, \sigma^2)$ for $\mu \in \mathbb{R}, \sigma^2 > 0$, if it has a p.d.f.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} \quad \text{for } x \in \mathbb{R}$$

$$\text{so} \quad \mathbb{E}[X] = \mu \quad \text{Var}(X) = \sigma^2$$

- For $Z \sim \mathcal{N}(0,1)$ the values of the c.d.f. can be found in a table

Distribution of a function of a r.v.

Distribution of a function of a r.v.

- If X is discrete or g is discrete,

$$\mathbb{P}(g(X) = k) = \mathbb{P}(X \in g^{-1}(\{k\}))$$

- If X and g are continuous, (e.g. g strictly increasing, invertible)
  1. Compute $\mathbb{P}(g(X) \leq t) = \mathbb{P}(X \leq g^{-1}(t)) = F_X(g^{-1}(t))$
  2. Deduce the p.d.f. of $Y = g(X)$ as

$$f_Y(y) = (g^{-1}(t))' f_X(g^{-1}(y))$$

- If g is not invertible, apply the above method on each interval on which g is invertible to compute the c.d.f. and deduce the p.d.f.

Law of large numbers

Law of Large Numbers

- For $X_1, \ldots, X_n$ i.i.d. r.v. with finite mean and variance, the **empirical mean** is
$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$
- The empirical mean "converges" to the true expectation $\mu = \mathbb{E}[X_i]$ as: for any $\varepsilon > 0$,
$$\lim_{n \to +\infty} \mathbb{P}(|\bar{X}_n - \mu| \leq \varepsilon) = 1$$

Normal Approximation

Normal approximation to Binomial

- If $S_n \sim \text{Bin}(n, p)$, consider the **standardization** of S_n, i.e.,

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - np}{\sqrt{np(1-p)}}$$

- Then the standardized S_n tends to be a standard normal dist. as $n \to +\infty$, i.e.

$$\lim_{n \to +\infty} \mathbb{P}\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

  this is called the **central limit theorem**

- Practically for $np(1-p) > 10$ (i.e. n large, p not too close to 0 or 1)

$$\mathbb{P}\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \approx \Phi(b) - \Phi(a)$$

  with Φ the c.d.f. of $Z\mathcal{N}(0, 1)$

Normal Approximation

Continuity correction

- When approximating for k_1, k_2 integers (or $k_1 = -\infty, k_2 = +\infty$),
$$\mathbb{P}(k_1 \leq S_n \leq k_2)$$
consider approximating
$$\mathbb{P}(k_1 - 1/2 \leq S_n \leq k_2 + 1/2)$$
This gives a better approximation when using the normal approx.

Confidence interval

- For $X \sim \text{Ber}(p)$ and $\hat{p} = S_n/n$ an estimate of p, we saw that
$$\mathbb{P}(|p - \hat{p}| \leq \varepsilon) \geq 2\Phi(2\varepsilon\sqrt{n}) - 1$$
- A confidence interval of level e.g. 95% consists of finding ε s.t.
$$\mathbb{P}(|p - \hat{p}| \leq \varepsilon) \geq 95\%$$
- To do that, compute $z_{95\%}$ s.t. $2\Phi(z_{95\%}) - 1 = 95\%$ (here $z_{95\%} = 1.96$)
- Then an $\varepsilon > 0$ s.t.
$$2\varepsilon\sqrt{n} \geq z_{95\%}$$
gives you a confidence interval!

Poisson Approximation

Poisson Approximation

- For $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Poisson}(np)$, for any $A \subset \{0, 1, 2, \ldots\}$,

$$|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \leq np^2$$

- So for $np^2 \ll 1$, the Poisson dist. is a better approximation of the binomial

Outline

p.d.f., c.d.f.

Exercise

Let X be a continuous random variable with density

$$f(x) = \begin{cases} cx^{-4} & \text{when } x \geq 1, \\ 0, & \text{elsewhere,} \end{cases}$$

where c is a constant.

1. What must be the value of c?
2. Find $\mathbb{P}(0.5 < X < 2)$.
3. Find the cumulative distribution function F_X.

Expectation, Variance, Quantiles

Exercise

Let X be a continuous random variable with density

$$f(x) = \begin{cases} 3e^{-3x} & \text{when } x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

1. Verify that f is a density function.
2. Find $\mathbb{E}[3X + 2]$.
3. Find $\text{Var}(2X + 8)$.
4. Find the 95% quantile of X.

Functions of r.v.

Exercise

Suppose X is $\mathrm{Exp}(\lambda)$.

1. Find the probability mass function of $Y = \lfloor X \rfloor$.
2. Find the probability density function of $Z = \log X$.

Approx. of binomials

Exercise

Every day John performs the following experiment. He flips a fair coin repeatedly until he sees a T and counts the number of flips needed.

1. Approximate the probability that in a year (365 days) there are at least 3 days when he needed more than 10 coin flips. Argue why this approximation is appropriate.
2. Approximate the probability that in a year (365 days) there are more than 50 days when he needed exactly 3 coin flips. Argue why this approximation is appropriate.