Recitation 13 Notes

Ziyang Lu, and Arick Grootveld

May 4th, 2022

Review For Exam

1 Foundations of Probability

- Use Venn Diagrams to describe:
  - $P(A \cap B)$: Intersection
  - $P(A \cup B)$: Union
  - $P(A^c)$: Complement
- **De Morgan's Law**:
  Steps to De Morgan's Law:
  1. Invert the individual elements
  2. Change all unions to intersections and all intersections to unions
  3. Invert the whole term

$$P(A \cap B) = P((A^c \cup B^c)^c)$$

- Other equations from this section:

$$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ P(A) &= 1 - P(A^c) \\ 0 \leq P(A) &\leq 1 \end{aligned}$$

2 Conditional Probability

- Conditional probability is used to describe the situation where something is known, and how a random variable changes based on this knowledge

- $P(A|B)$ can be read in plain English as: The probability of the event A given that the event B happened

- **Equations from this section**:

$$\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
P(A) &= P(A|B)P(B) + P(A|B^c)P(B^c) \\
P(A|B) &= \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \\
P(A^c|B) &= 1 - P(A|B) \\
P(A^c|B^c) &= 1 - P(A|B^c)
\end{aligned}$$

3 Discrete Random Variables

- **Probability mass function**

  If X is a Discrete Random Variable (DRV), then it has some probability mass function (pmf) such that $f_X(x) = P(X = x)$

- **Cumulative Distribution Function**

  If X is a DRV, with pmf $f_X(x)$, then it also has a Cumulative Distribution Function (cdf), $F_X(x) = P(X \leq x)$

  Note that sometimes the cdf of a RV will just be called the "distribution function"

- **Probability Intervals**:

  If X is a DRV with pmf $f_X(x)$, then:

$$P(a \leq X \leq b) = \sum_{i=a}^{b} f_X(x)$$

- **Geometric Series**

  While the geometric series is not directly a part of discrete random variables, it is heavily used to calculate the cdf or expected value of some discrete random variables

  - *Infinite Geometric series*:

$$\begin{aligned}
1 + x + x^2 + \cdots &= \sum_{i=0}^{\infty} x^i \\
&= \frac{1}{1-x}, \text{ for } 0 < x < 1
\end{aligned}$$

- *Finite Geometric Series*:

$$\begin{aligned} 1 + x + x^2 + \cdots + x^n &= \sum_{i=0}^{n} x^i \\ &= \frac{1 - x^{n+1}}{1 - x}, \text{ for } 0 < x < 1 \end{aligned}$$

- Law of total probability applied to DRVs
  If X is a DRV with pmf $f_X(x)$ and cdf $F_X(x)$, then

$$P(-\infty < X < \infty) = \sum_{i=-\infty}^{\infty} f_X(i) = 1$$

$$F_X(\infty) = P(X \leq \infty) = \sum_{i=-\infty}^{\infty} f_X(i) = 1$$

- **Common Discrete Distributions**
  - *Bernoulli Distribution* $\{\text{bern}(p)\}$:
    The Bernoulli distribution represents flipping an unfair coin that has probability p of coming up heads, and counting a heads as a 1, and tails as a 0.
    If $X \sim \text{bern}(p)$

$$\begin{aligned} f_X(x) &= (p)^x (1-p)^{1-x} \\ &\text{For } x \in \{0, 1\} \\ E[X] &= p \\ \text{Var}(X) &= p(1-p) \end{aligned}$$

  - *Binomial Distribution* $\{\text{bin}(n, p)\}$
    The binomial distribution is a sum of Bernoulli random trials
    If $X \sim$ bin(n,p), then:

$$\begin{aligned} f_X(k) &= \binom{n}{k} p^k (1-p)^{n-k}, \text{ for } k = 0, 1, 2, \cdots, n \\ E[X] &= n \cdot p \\ \text{Var}(X) &= n \cdot p \cdot (1-p) \end{aligned}$$

  - *Geometric Distribution* $\{\text{Geo}(p)\}$ A geometric distribution can be described as the number of Bernoulli trials required to get a success
    If $X \sim$ Geo(p), then:

$$\begin{aligned} f_X(k) &= (1-p)^{k-1} \cdot p \\ E[X] &= \frac{1}{p} \\ \text{Var}(X) &= \frac{1-p}{p^2} \end{aligned}$$

4 Continuous Random Variables

- The major difference between a DRV and a continuous random variable (CRV) is that we switch from interpreting probabilities as sums of finite (or countably infinite) points, to areas under the curve.
  - For example: If X is a CRV, then we can always say $P(X = c) = 0$, where c is any constant
- **Probability Density Function** If X is a CRV, then it has a probability density function (pdf) $f_X(x)$, such that

$$P(a \leq X \leq b) = \int_a^b f_X(x)\, dx$$

- When talking about CRV's, we note that the probability of a CRV being exactly equal to any particular number is 0

  This also means that

$$\begin{aligned} P(a \leq X \leq b) &= P(X = a) + P(a < X < b) + P(X = b) \\ &= 0 + P(a < X < b) + 0 \\ \implies P(a \leq X \leq b) &= P(a < X < b) \end{aligned}$$

  So we can exchange inclusive inequalities ($\leq, \geq$) with exclusive inequalities ($<, >$) as we want when we are talking about CRVs
- **Cumulative Distribution Function**

  One thing that DRVs and CRVs have in common is that they both have a cdf

  So if X is a CRV with pdf $f_X(x)$, then X also has a cdf $F_X(x)$, such that:

$$F_X(a) = P(X \leq a) = \int_{-\infty}^{a} f_X(x)\, dx$$

- Law of total probability applied to CRVs

  In the same vein as DRVs, if X is a CRV with pdf $f_X(x)$ and cdf $F_X(x)$, then

$$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

$$F_X(\infty) = P(X \leq \infty) = \int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

- **Common Continuous Distributions**:

- *Uniform Distribution* $\{\text{Unif}(a,b)\}$

  This distribution describes the event where the random variable is some value inside the interval $[a,b]$, and it can take on any of the values in that interval with the same probability

  If $X \sim \text{Unif}(a,b)$, then

$$
\begin{aligned}
f_X(x) &= \frac{1}{b-a}, \text{ for } a \le x \le b \\
E[X] &= \frac{a+b}{2} \\
\text{Var}(X) &= \frac{(b-a)^2}{12}
\end{aligned}
$$

- *Exponential Distribution* $\{\text{Exp}(\lambda)\}$

  If $X \sim \text{Exp}(\lambda)$ for $\lambda > 0$, then

$$
\begin{aligned}
f_X(x) &= \lambda e^{-\lambda x} \\
E[X] &= \frac{1}{\lambda} \\
\text{Var}(X) &= \frac{1}{\lambda^2}
\end{aligned}
$$

- *Normal Distribution* $\{\text{N}(\mu,\sigma^2)\}$

  If $X \sim \text{N}(\mu,\sigma^2)$, then

$$
\begin{aligned}
f_X(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\
E[X] &= \mu \\
\text{Var}(X) &= \sigma^2
\end{aligned}
$$

5 Joint distribution

1) Discrete R.V:

1. Joint pmf: $P(X=a, Y=b)$ or $P_{XY}(a,b)$.
2. Marginalization: $P_{XY} --> P_X$ and P_Y?

$$P_X(a) = \sum_b P_{XY}(a,b)$$

$$P_Y(b) = \sum_a P_{XY}(a,b)$$

   (sum rows and columns in the table). **See quiz 5, problem 2**
3. Joint cdf: $F_{XY}(a,b) = P(X \le a, Y \le b)$

4. $F_{XY} --> F_X, F_Y$?
   $F_X(a) = P(X \leq a) = F_{XY}(a, \infty)$ (Let $b = \infty$)
   $F_Y(b) = P(Y \leq b) = F_{XY}(\infty, b)$ (Let $a = \infty$)
   See quiz 5, problem 1

2) Continuous R.V.

1. Joint pdf: $f(x, y)$.
2. Marginalization: $f(x, y) --> f(x)$ and $f(y)$?

$$f(x) = \int f(x, y)dy$$

$$f(y) = \int f(x, y)dx$$

   See hw5, problem 6

3. Given the joint pdf f(x, y), how to compute probability?
   (e.g. find $P(a \leq X \leq b, c \leq Y \leq d)$)

$$P(a \leq X \leq b, c \leq Y \leq d) = \int \int f(x, y)dydx$$

   Note that the ranges of variables are not filled in these equations. It is the critical part of solving problems in continuous R.V.

   See quiz 5 problem 3(easy)

   HW5 problem 4 (normal)

   HW5 problem 6 (hard)

3) Independence:
If X, Y are independent,

1. Discrete R.V.: $P(X = a, Y = b) = P(X = a)P(Y = b)$
2. Continuous R.V.: $f(x, y) = f(x)f(y)$ ($f(x)$ and $f(y)$ can be found by marginalization)

6 Covariance (Covered in Recitation 12)

- 1. $Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$
- 2. $Cov(X, X) = Var(X)$
- 3. $Cov(X, Y) = Cov(Y, X)$
- 4. $Cov(X, c) = E[(X - E(X))(c - E(c))] = 0$

- 5. $Cov(X, Y + Z) = Cov(X, Y) + Cov(X + Z)$

  $Cov(X + Y, Z + K) = Cov(X, Z) + Cov(X, K) + Cov(Y, Z) + Cov(Y, K)$ (Bilinearily property)

- 6. $Cov(X, Y) = E(XY) - E(X)E(Y)$

- 7. If two R.V. X and Y are independent Cov$(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$

  But given $Cov(X, Y) = 0$, we cannot say X and Y are independent. (One way)

  See hw6 problems

6.1 Correlation $\rho(X, Y)$ (Covered in Recitation 12)

$\rho(X, Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$. We have $-1 \leq \rho \leq 1$

See hw6 problem 1

7 Graphical Summaries

1). Height of a bin:

If we have a continuous random variable X. We can sample a bunch of points (x_i) to estimate the probability.

$$height_{(a,b]} = \frac{P(a < X \leq b)}{b - a}$$

See hw6 problem 3 2). Histogram: A graphical representation of the heights of bins. Compute the heights and draw it.

8 Numerical Summaries

1). Sample mean
2). Sample median
3). Sample variance
4). Sample standard deviation
5). MAD (mean absolute deviation)
Example:
Samples: 1, 5, 3, 3, 1, 3

1) Sample mean: $E(X) = \frac{1+5+3+3+1+3}{6} = 16/6 = 8/3$

2) Sample median: We first sort the sequence: 1, 1, 3, 3, 3, 5, then the median is $med = \frac{3+3}{2} = 3$

3) Sample variance:

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - E(X))^2$$

$$= \frac{1}{6-1}[(1-\frac{8}{3})^2 + (5-\frac{8}{3})^2 + (3-\frac{8}{3})^2 + (3-\frac{8}{3})^2 + (1-\frac{8}{3})^2 + (3-\frac{8}{3})^2] =$$

4) Sample standard deviation: $s = \sqrt{s^2}$

5) MAD:

$$MAD = med(|1-med|, |5-med|, |3-med|, |3-med|, |1-med|, |3-med|)$$

$$= med(2, 2, 0, 0, 2, 0) = \frac{0+2}{2} = 1$$

Sort the sequence before finding median.
See hw6 problems.