STAT-4300 - Class Notes

April 29, 2024

Chapter 1: Random Variables & Expectations

1.1 Function of random variables

Given sample space S, r.v. X, and function $g : \mathbb{R} \to \mathbb{R}$, $g(X)$ is the r.v. that takes value $g(X(s))$ for any outcome $s \in S$

If X, Y are r.v., any function of $g(X, Y)$ is also a r.v.

1.2 Independence of random variables

Definition: X, Y independent if

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y) \text{ for all } x, y \in \mathbb{R}$$

$X_1, X_2, \ldots X_n$ are independent if

$$P(X_1 \leq x_1, \ldots X_n \leq x_n) = P(X_1 \leq x_1) \cdot \ldots \cdot P(X_n \leq x_n)$$

Theorem: If X, Y independent, any function of X is independent of any function of Y

Theorem: IF X, Y are discrete, the following are equivalent

- X, Y independent
- $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$
- $P(X = x, Y = y) = P(X = x)P(Y = y)$
- $P(X = x|Y = y) = P(X = x)$
- $P(Y = y|X = x) = P(Y = y)$

Definition $X_1, \ldots X_n$ are independent and identically distributed if

- $X_1 \ldots X_n$ independent
- $X_1 \ldots X_n$ have the same distribution

Definition: X, Y are conditionally independent given Z if

$$P(X \leq x, Y \leq y|Z \leq z) = P(X \leq x|Z \leq z)P(Y \leq y|Z \leq z)$$

Definition: The conditional PMF of X given Z is

$$P(X = x|Z = z)$$

1.3 Expectation

Definition: The expected value of a discrete r.v. X whose possible values are $x_1, x_2 \dots$ is

$$E[X] = \sum_{j=1}^{\infty} x_j P(X = x_j)$$

1.3.1 Linearity of expectation

For any r.v. X, Y and constant c

- $E[X + Y] = E[X] + E[Y]$
- $E[cX] = cE[X]$

Chapter 2: Lesson 11

2.1 Geometric distribution

For a sequence of of Bernoulli trials, each with the same success probability p, the number of trials **before** first success is a **geometric distribution with parameter** p

$$X \sim Geom(p)$$

$$X + 1 \sim FS(p)$$

PMF:

$$P(X = k) = (1 - p)^k p$$

Expectation

$$\begin{aligned} E[X] &= \sum_{k=0}^{\infty} k \cdot P(K = k) \\ &= \sum_{k=0}^{\infty} k(1-p)^k p \\ &= p \sum_{k=0}^{\infty} kq^k \\ &= p \frac{q}{(1-q)^2} \\ &- \frac{1-p}{p} \end{aligned}$$

$$\begin{aligned} E[FS(p)] &= E[X + 1] \\ &= E[X] + 1 \\ &= \frac{1}{p} \end{aligned}$$

2.2 Indicator RVs

Definition: The indicator r.v. for an event A is the r.v. I_A that takes value 1 if A occurs, and 0 if A does not occur

$$I_A \sim Ber(p) \text{ where } p = P(A)$$

Theorem: Fundamental bridge between probability and expectation

$$E[I_A] = p = P(A)$$

Properties

1. $I_{A^C} = 1 - I_A$
2. $I_{A \cap B} = I_A \cdot I_B$
3. $I_{A \cup B} = I_A + I_B - I_A \cdot I_B$

2.3 Darth Vader Rule - Expectation via Survival Function

Theorem: Let X be a discrete r.v. with support $[0, 1, \ldots]$

$$E[X] = \sum_{n=0}^{\infty} P(X > n) = \sum_{n=1}^{\infty} P(X \geq n)$$

Where

$$\begin{aligned} P(X > n) &= 1 - P(X \leq n) \\ &= 1 - F(n) \text{ known as the survival function} \end{aligned}$$

For X with support $[1 \ldots N]$

$$E[X] = \sum_{n=0}^{N} P(X > n) = \sum_{n=1}^{N} P(X \geq n)$$

Reasoning

$$\begin{aligned} X &= \sum_{n=1}^{N} I_n \text{ where} \\ I_n &= \mathbb{1}[X \geq n] \end{aligned}$$

Chapter 3: Moment Generating Functions

3.1 Moments

Definition: For any r.v..X and any $n = 1, 2, 3 \ldots$

- n-th moment:
$$E[X^n]$$
- n-th central moment:
$$E[(X-\mu)^n]$$
- n-th standardized moment:
$$E\left[(\frac{X-\mu}{\sigma})^n\right]$$

Motivation: Moments are often used as summaries of a distribution

- Mean
- Variance
- Skewness
- Excess kurtosis

3.2 Moment Generating Functions

Definition: The MGM of a r.v. X is a function of t

$$M_X(t) = E\left[e^{tX}\right]$$

Note that MGM is a function from $\mathbb{R}$ to $\mathbb{R}$
The MGM might be for some values of t, we say the MGM exists if it is finite for all t within some open interval

3.2.1 Moment Generating Function of Bernoulli

3.2.2 Moment Generating Function of Uniform

3.2.3 Moment Generating Function of Exponential

Given $X \sim Expo(1)$

$$\begin{aligned} M_X(t) &= E\left[e^{tX}\right] \\ &= \int_0^\infty e^{tx} \end{aligned}$$

3.3 Why MGFs are useful

Theorem: Derivatives of MGFs give the moments

If the MGF exists

$$E[X^n] = M^{(n)}(0) \text{ for } n = 1, 2, \ldots$$

Theorem: MGF determines the distribution
If X, Y have the same MGF then they have the same distribution

Theorem: MGF of independent r.v.s

If X and Y are independent and their MGFs exist, then

$$MGF_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$

Chapter 4: Class 21

4.1 Covariance

Definition: For r.v. X, Y, the **covariance** is defined as

$$Cov(X,Y) = E\left[(X - E[X]) \cdot (Y - E[Y])\right]$$

Equivalently,

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

Note that

$$Var(X) = Cov(X,X)$$

Properties of covariance

- Independent r.v.s have zero covariance

$$X \perp Y \implies Cov(X,Y)$$

- Symmetry

$$Cov(X,Y) = Cov(Y,X)$$

- Covariance between r.v. and constant is zero

$$Cov(X,c) = 0 \text{ for any constant } c$$

- Scaling by a constant

$$Cov(aX,Y) = aCov(X,Y)$$

- Sum of two random variables

$$Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$$

- Covariance of sum of r.v.s

$$Cov(X_1 + X_2, Y_1 + Y_2) = Cov(X_1, Y_1) + Cov(X_1, Y_2) + Cov(X_2, Y_1) + Cov(X_2, Y_2)$$

- More generally,

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

Proofs of properties

Proof of (1):

$$\begin{aligned} X \perp Y &\implies E[XY] = E[X]E[Y] \\ &\implies Cov(X,Y) = E[XY] - E[X]E[Y] = E[XY] - E[XY] = 0 \end{aligned}$$

Proof of (2)

$$\begin{aligned}
Cov(X,Y) &= E[XY] - E[X]E[Y] \\
&= E[YX] - E[Y]E[X] \\
&= Cov(Y,X)
\end{aligned}$$

Proof of (3)

$$\begin{aligned}
Cov(X,c) &= E[cX] - E[c]E[X] \\
&= cE[X] - cE[X] \\
&= 0
\end{aligned}$$

Proof of (4)

$$\begin{aligned}
Cov(aX,Y) &= E[aXY] - E[aX]E[Y] \\
&= aE[XY] - aE[X]E[Y] \\
&= a\left(E[XY] - E[X]E[Y]\right) \\
&= aCov(X,Y)
\end{aligned}$$

Proof of (5)

$$\begin{aligned}
Cov(X_1+X_2,Y) &= E[(X_1+X_2)Y] - E[X_1+X_2]E[Y] \\
&= E[X_1Y + X_2Y] - E[X_1+X_2]E[Y] \\
&= E[X_1Y] + E[X_2Y] - E[X_1]E[Y] - E[X_2]E[Y] \\
&= E[X_1Y] - E[X_1]E[Y] + E[X_2Y] - E[X_2]E[Y] \\
&= Cov(X_1,Y) + Cov(X_2,Y)
\end{aligned}$$

Proof of (6)

$$\begin{aligned}
&Cov(X_1+X_2,Y_1+Y_2) \\
=&Cov(X_1,Y_1+Y_2) + Cov(X_2,Y_1+Y_2) \quad \text{by Property 5} \\
=&Cov(X_1,Y_1) + Cov(X_1,Y_2) + Cov(X_2,Y_1) + Cov(X_2,Y_2)
\end{aligned}$$

4.2 General case for variance of a sum of r.v.s

Theorem: For any X,Y

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

$$Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$$

Proof

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

OR

$$\begin{aligned}
&Cov(X+Y,X+Y) \\
=&Cov(X,Y) + Cov(X,Y) + Cov(Y,X) + Cov(Y,Y) \\
=&Var(X) + 2Cov(X,Y) + Var(Y)
\end{aligned}$$

4.2.1 Example

Let $X, Y \sim Unif(0,1)$, find $Cov(X+Y, X-Y)$

$$\begin{aligned} Cov(X+Y, X-Y) &= Var(X) - Var(Y) \\ &= 0 \end{aligned}$$

$X+Y, X-Y$ not independent

$$X+Y=2 \iff X=Y=1 \iff X-Y=0$$

4.3 Correlation

*not on Finals

Definition: Correlation between X, Y is

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y}} = \frac{Cov(X,Y)}{sd(X) \cdot sd(Y)}$$

*Correlation is the rescaled version of $Cov(X,Y)$

Properties

- Location-scale transforms have no effect on correlations

$$Corr(aX+b, cY+d) = Corr(X,Y) \text{ for any } a>0, c>0$$

- Correlation doesn't depend on units of measurement
- Correlation takes values from -1 to 1

$$-1 \leq Corr(X,Y) \leq 1$$

- Perfect correlation implies perfect linear relationship

$$|Corr(X,Y)| = 1 \implies Y = aX+b$$

4.4 Multivariate Gaussian Distribution

*not on Finals

Definition: a **random vector** is an ordered list of random variables

Definition: a random vector $(X_1 \dots X_k)$ has the Multivariate Gaussian distribution if every linear combination of the X_j s has a Gaussian distribution.

$$(X_1 \dots X_k) \sim MVG \text{ if } \forall \mathbf{t} \in \mathbb{R}^k$$
$$\left(\sum_{i=1}^{k} t_i X_i\right) \sim N(\mu, \sigma^2) \text{ for some } \mu, \sigma$$

Chapter 5: Class 22 - Transformations

5.1 Discrete random variables

Discrete: Given the distribution of X, to compute the distribution of $Y = g(X)$

$$P(Y = y) = P(g(X) = y) = \sum_{x s.t. g(x)=y} P(X = x)$$

If g invertible,

$$P(Y = y) = P(g(X) = y) = P(X = g^{-1}(y))$$

5.2 Continuous random variables

5.2.1 Via CDF

Continuous:

Method 1

Step 1: Find the CDF of Y

For order-preserving, invertible g,

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y) = F_X(g^{-1}(y))$$

Step 2: differentiate to get the PDF of Y

5.2.2 Via change in variable

Theorem: Suppose X is a continuous r.v. with PDF f_X, let $Y = g(X)$ where g is differentiable, and strictly increasing or decreasing on the support of X, then the PDF of y is

$$f_Y(y) = f_X(x) \cdot \left|\frac{dx}{dy}\right| \text{ where } x = g^{-1}(y)$$

Note

- $$\frac{dx}{dy} = \frac{1}{dy/dx}$$
- Support of Y is the range of g whose input is restricted to the support of X

$$\mathit{support}(Y) = \{y = g(x) \text{ for some } x \in \mathit{support}(X)\}$$

In multivariable setting

Given $\vec{Y} = (Y_1, Y_2 \dots Y_n)$, $\vec{y} = (y_1, y_2 \dots y_n)$,

$$f_{\vec{Y}}(\vec{y}) = f_{\vec{X}}(\vec{x}) \cdot \left| \frac{\partial \vec{x}}{\partial \vec{y}} \right|$$

Chapter 6: Convolutions and Conditional expectation

6.1 Convolutions

Theorem [Discrete] Let X, Y be independent r.v.s, the PMF of $T = X + Y$ is

$$P(T = t) = \sum_x P(Y = t - x) \cdot P(X = x)$$

Justification: Law of Total Probability

$$\text{LOTP} : P(B) = \sum_i P(B|A_i)P(A_i)$$

In this case

$$\begin{aligned} P(T = t) &= P(X + Y = t) \text{ by definition} \\ &= \sum_x P(X + Y = t|X = x) \cdot P(X = x) \text{ LOTP} \\ &= \sum_x P(Y = t - x|X = x) \cdot P(X = x) \text{ by rearranging} \\ &= \sum_x P(Y = t - x) \cdot P(X = x) \text{ by independence of } X, Y \end{aligned}$$

Theorem [Continuous] Let X, Y be independent r.v.s, the PDF of $T = X + Y$ is

$$f_T(t) = \int_{\infty}^{\infty} f_X(x) \cdot f_Y(t - x) dx$$

6.2 Conditional expectation

Definition: Let A be an event with $P(A) > 0$

Y is **discrete**

$$E[Y|A] = \sum_y yP(Y|A)$$

Y is **continuous**

$$E[Y|A] = \int_{\infty}^{\infty} yf(Y|A)dy$$

Chapter 7: Conditional Expectation

7.1 Definitions

Definition Conditional given event

- discrete Y

$$E[Y|A] = \sum_y yP(Y = y|A)$$

- continuous Y

$$E[Y|A] = \int_{\infty}^{\infty} yf_Y(y|A)dy$$

Definition Conditional given the value of a random variable X

- If X is discrete

$$E[Y|X = x] = E[Y|A] \text{ where } A \text{ is the event that } X = x$$

- if X, Y continuous

$$E[Y|X = x] = \int_{\infty}^{\infty} yf_{Y|X}(y|x)dy$$

- if X continuous, Y discrete

$$E[Y|X = x] = \sum_y yP(Y = y|X = x)$$

Definition: Conditional given another random variable

Let X, Y be r.v.s, let g be function

$$g : val(X) \to val(E[Y|X])$$

The **conditional expectation** of Y, given X is the r.v. $g(X)$

$$E[Y|X] = g(X)$$

7.2 Properties of conditional expectation

Ignoring what's independent

$$X \perp Y \to E[Y|X] = E[Y]$$

Taking out what's known:

$$E[h(X)Y|X] = h(X) \cdot E[Y|X]$$

Linearity

$$E[Y_1 + Y_2|X] = E[Y_1|X] + E[Y_2|X]$$

Law of Iterated Expectation / Law of Total Expectations / Adam's Law / Tower Law

$$E[Y] = E\left[[Y|X]\right]$$