# Stat 88: Probability & Mathematical Statistics in Data Science


Lecture 13: 2/19/2021

Functions of rvs, Method of indicators

Sections 5.2, 5.3

$$\mathbb{E}(X) = \sum_x x \cdot P(X=x) = \sum_x x \cdot f(x)$$

## Special examples

Bernoulli(p) $\Longleftrightarrow$ Bin(1, p)

- Bernoulli (Indicators) $X \sim$ Bernoulli(p)

Indicator variables, for any event $A$, $P(A) = p > 0$ define an associated r.v. $I_A = \begin{cases} 1, & A \text{ is true} \\ 0, & A \text{ is false} \end{cases}$

$$\mathbb{E}(I_A) = 1 \cdot P(A) + 0 \cdot P(A^c) = P(A)$$

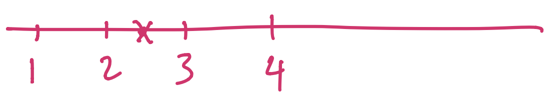
- Uniform $X \sim$ Unif on $\{1, 2, \ldots n\}$

$$E(X) = \sum_x x \cdot P(X=x) \qquad P(X=k) = \frac{1}{n}$$

$$= 1 \cdot \frac{1}{n} + 2 \cdot \frac{1}{n} + \cdots + n \cdot \frac{1}{n} = \frac{1}{\cancel{n}} [\underbrace{1+2+3+\cdots+n}_{\frac{\cancel{n}(n+1)}{2}}]$$

- ~~Poisson~~

$$\mathbb{E}(X) = \frac{1}{n} \cdot \left[ n \frac{(n+1)}{2} \right] = \frac{n+1}{2}$$





$$X = \begin{cases} 0 & \text{w/prob } (1-p) \\ 1 & \text{w/prob } p. \end{cases}$$

Boxes

$$X = \begin{cases} 0 & \text{w.p } 2/3 \\ 1 & \text{w.p. } 1/3 \end{cases}$$

$$\mathbb{E}(X) = 0 \cdot \frac{2}{3} + 1 \cdot \frac{1}{3} = \frac{1}{3} \longrightarrow \boxed{0}\ \boxed{0}\ \boxed{1}$$

$$\frac{0+1+2+3}{4} \qquad X = \begin{cases} 0 \\ 1 \\ 2 \\ 3 \end{cases} \text{w/p } \frac{1}{4}$$

$$E(X) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4}$$

$$\longrightarrow \boxed{0}\ \boxed{1}\ \boxed{2}\ \boxed{3}$$

$$X = \begin{cases} 0 & \text{w.p } 3/10 \\ 1 & \text{w.p } 1/2 \\ 2 & \text{w.p. } 2/10 \end{cases}$$

$$\mathbb{E}(X) = 0 \cdot \frac{3}{10} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{2}{10} = 1 \cdot \frac{5}{10} + 2 \cdot \frac{2}{10} = \frac{9}{10}$$

$$\underbrace{\boxed{0}\boxed{0}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{1}\boxed{1}\boxed{2}\boxed{2}}_{\text{avg} = \frac{9}{10}.}$$


middle is balancing point $= \frac{n+1}{2}$ ← when X is uniform.

## Poisson

$$P(X=k) = e^{-\mu} \cdot \frac{\mu^k}{k!}$$

$$\mathbb{E}(X) = \sum_x x \cdot P(X=x) = \sum_{k=0}^{\infty} k \cdot e^{-\mu} \frac{\mu^k}{k!}$$

$$= e^{-\mu} \sum_{k=0}^{\infty} \cancel{k} \cdot \frac{\mu^k}{\cancel{k!}_{(k-1)!}} = e^{-\mu} \underbrace{\sum_{k=1}^{\infty} \frac{\mu^k}{(k-1)!}}_{} \Big\} \, e^{\mu}$$

$k \cdot \frac{\mu^k}{k!}$ at $k=0$, this is 0

Recall from Calc.

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots = \sum_{j=0}^{\infty} \frac{x^j}{j!}$$

$$\mathbb{E}(X) = e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^k}{(k-1)!} = e^{-\mu} \mu \sum_{j=0}^{\infty} \frac{\mu^j}{j!}$$

$j = k-1$

$$= e^{-\mu} \cdot \mu \cdot e^{\mu} = \mu$$

# 5.2: Functions of random variables

$$E(X) = \sum_x x \cdot P(X=x)$$

$$Y = g(X)$$

- X~unif{-1,0,1}, Y=X², find E(Y)

| $x$ | $-1$ | $0$ | $1$ |
|---|---|---|---|
| $f(x)$ | $1/3$ | $1/3$ | $1/3$ |
| $y$ | $(-1)^2=1$ | $0^2=0$ | $1^2=1$ |

$$\sum_x x^2 P(X=x)$$

$$E(Y) = 1\left(\frac{1}{3}\right) + 0\cdot\left(\frac{1}{3}\right) + 1\left(\frac{1}{3}\right) = 1\left(\frac{2}{3}\right) + 0 = 2/3$$

- In general, if $Y = g(X), E(Y) = E\big(g(X)\big) = \sum g(x) \cdot f(x) = \sum g(x) \cdot P(X = x)$

$$Y = g(X), \quad E(Y) = E(g(X)) = \sum_x g(x) \cdot P(X=x)$$

- Example: $W = \min(X, 0.5)$. Find E(W) (write out the values of W, and probs)

| $x$ | $-1$ | $0$ | $1$ |
|---|---|---|---|
| $W$ | $-1$ | $0$ | $0.5$ |
| $f(x)$ | $1/3$ | $1/3$ | $1/3$ |

$$E(W) = (-1)\left(\frac{1}{3}\right) + 0\left(\frac{1}{3}\right) + (0.5)\left(\frac{1}{3}\right) = -\frac{1}{3} + \frac{0.5}{3} = -\frac{0.5}{3}$$

- Note that $Y = g(X)$ is also a random variable, just defined via X.

$$E(Y) = a \cdot E(X) + b$$

$$Y = aX + b$$

$$g(x) = ax + b$$

$$E(Y) = \sum_x g(x) f(x) = \sum_x (ax+b) f(x) = \sum ax \cdot f(x) + \sum_x b \cdot f(x)$$

$$= a \underbrace{\sum_x x \cdot f(x)}_{a\,\mathbb{E}(X)} + b \underbrace{\sum f(x)}_{b \cdot 1}$$

# Multiple random variables on the same outcome space

- Joint distributions: Recall rolling a pair of dice. Draw a table of outcomes and their probabilities. What event does each cell represent?
- Sum of probabilities across *all* the cells is 1, since this is all the probabilities across all possible outcomes

| $X_1$ \ $X_2$ | 1 | 2 | 3 | |
|---|---|---|---|---|
| 1 | (1, 1) | (1, 2) | (1, 3) | 1/3 |
| 2 | (2, 1) | (2, 2) | (2, 3) | |
| 3 | (3, 1) | (3, 2) | (3, 3) | |

Prob: 1/9 for each

- Now, suppose we draw two tickets without replacement from a box that has 5 tickets marked 1, 2, 2, 3, 3. Let $X_1$ and $X_2$ represent the values of the tickets drawn on the first and second draws respectively.
- Create a table of all possible outcomes for the pair ($X_1$, $X_2$) (which is also a random variable), and write down the probabilities using the multiplication rule.

# Joint distributions

[1] [2] [2] [3] [3]

- Draw two tickets without replacement from a box that has 5 tickets marked 1, 2, 2, 3, 3. Let $X_1$ and $X_2$ represent the values of the tickets drawn on the first and second draws respectively.

$5 \cdot 4 = 20$

- Create a table of all possible outcomes for the pair ($X_1$, $X_2$) (which is also a random variable), and write down the probabilities using the multiplication rule.

Joint distribution of $X_1, X_2$

| | $X_2 = 1$ | $X_2 = 2$ | $X_2 = 3$ | |
|---|---|---|---|---|
| $X_1 = 1$ | $0 = P(X_1=1, X_2=1)$ | $\frac{2}{20} = P(X_1=1, X_2=2)$ | $\frac{2}{20} = P(X_1=1, X_2=3)$ | $0 + \frac{2}{20} + \frac{2}{20} = \frac{4}{20} = \frac{1}{5}$ |
| $X_1 = 2$ | $\frac{2}{20}$ | $\frac{2}{20} = P(X_1=2, X_2=2)$ | $\frac{4}{20}$ | $\frac{2}{20} + \frac{2}{20} + \frac{4}{20} = \frac{8}{20}$ |
| $X_1 = 3$ | $\frac{2}{20} = P(X_1=3, X_2=1)$ | $\frac{4}{20} = P(X_1=3, X_2=2)$ | $\frac{2}{20}$ | $\frac{8}{20}$ |
| | $4/20$ | $8/20$ | $8/20$ | |

$f(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$ = Joint p.m.f for $X_1$ & $X_2$.

# Marginal distributions

$X_1 = \begin{cases} 1 & \text{w/prob } \frac{4}{20} = \frac{1}{5} \\ 2 & \text{w/prob } 8/20 \\ 3 & \text{w/p } 8/20 \end{cases}$

$X_2 = \begin{cases} 1 & \text{wp } 4/20 \\ 2 & \text{wp } 8/20 \\ 3 & \text{wp } \frac{8}{20} \end{cases}$

- What is $\mathrm{P}(X_1 = 1)$? Write down the pmf for $X_1$ and $X_2$
- Are they independent? Not indep
- Use the table to compute $P(X_1 + X_2 = 5)$ $= \frac{4}{20} + \frac{4}{20} = \frac{8}{20}$
- Use the table to compute $E\big(g(X_1, X_2)\big)$, where $g(X_1, X_2) = | X_1 - X_2 |$ (the expected distance between the two draws)

$= g(1,1)\cdot 0 + g(1,2)\cdot \frac{2}{20} + g(1,3)\cdot\frac{2}{20}$
$+ \ldots$
(9 terms in this addition) $= 1$

| $X_1 \backslash X_2$ | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0 | 1 | 2 |
| 2 | 1 | 0 | 1 |
| 3 | 2 | 1 | 0 |

$g(X_1, X_2)$

$(X_1, X_2)$

- If $S = X_1 + X\_2$, find $E(S)$.

Do the same.

& verify that $= 1$ ... $= EX + EX_2$

| $X_1 \backslash X_2$ | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 2 | 3 | 4 | 5 |
| 3 | 4 | 5 | 6 |

$S = X_1 + X_2$

$g(X_1, X_2)$

Probs.

| $X_1 \backslash X_2$ | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0 | 2/20 | 2/20 |
| 2 | 2/20 | 2/20 | 4/20 |
| 3 | 2/20 | 4/20 | 2/20 |

$P(X_1=1, X_2=3)$

[illegible] $X_2)$ [illegible] $g$ [illegible] )