Lecture 5 - 8/31

Review:

- Random variable: function from Ω to $\mathbb{R}$
  - realization of a random variable
  - discrete random variable
  - pmf: $p_X(u) = P(X=u) = \mathbb{P}(\{\omega \in \Omega \text{ s.t. } X(\omega)=u\})$

    satisfies axioms of probability.
  - support of random variable
  - function of random variable $Y = g(X)$

$$p_Y(u) = P(Y=u) = \sum_{x: g(x)=u} p_X(u)$$

- Expected value, mean, average
  - $E[X] = \sum_u u\, p_X(u)$
  - LOTUS: $E[g(X)] = \sum_u g(u) p_X(u)$
- Frequency interpretation:

  Suppose we ran an experiment N times and the outcome of each trial is $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$
  - Suppose that ω_i is observed M_i times as an outcome
    - frequency of observing $\omega_i = \frac{M_i}{N} \approx \mathbb{P}(\{\omega_i\})$
  - Suppose we have a random variable $X: \Omega \longrightarrow \{x_1, x_2, \dots, x_n\}$

    passing outcomes of experiment through X, suppose passing outcomes of experiment through X, suppose that x_i is observed N_i times.

- frequency of observing $x_i = \frac{N_i}{N} \approx P_X(x_i)$
- average number observed after the experiment:

$$\frac{x_1 . N_1 + x_2 . N_2 + \cdots + x_n . N_n}{N} \approx x_1 \cdot P_X(x_1) + \cdots + x_n \cdot P_X(x_n) = E[X]$$

Today:

① LOTUS, Variance, moments

② Conditional probability

---

① LOTUS, Variance, moments

(1.1). LOTUS: let $Y = g(X)$

$$E[Y] = \sum_u u\, P_Y(u) = \sum_u u \sum_{v: g(v)=u} P_X(v)$$

$$= \sum_{u,v: g(v)=u} g(v) P_X(v) = \sum_v g(v) P_X(v)$$

i.e., $E[g(X)] = \sum_u g(u) P_X(u)$

. LOTUS implies:

$$E[a g(X) + b h(X) + c] = a E[g(X)] + b E[h(X)] + c$$

(1.2). Variance and standard deviation

Def. Variance of a random variable X measure how spread out

Def. Variance of a random variable X measure how spread out the pmf of X is:

$$Var(X) = E[(X-\mu_X)^2]$$

where

$$\mu_X = E[X]$$

- $X-\mu_X$ is deviation from mean, the error if we predict X with μ_X. Notice that $E[X-\mu_X]=0$
- Why not $|X-\mu_X|$? since $(X-\mu_X)^2$ is differentiable.

Def. Standard deviation of X:

$$\sigma_X = \sqrt{Var(X)}$$

- by LOTUS:
  - $Var(X) = E[(X-\mu_X)^2] = E[X^2+\mu_X^2-2\mu_X X]$

    $= E[X^2] - \mu_X^2$: simplify & see why!
  - $E[aX] = a\mu_X$

    $Var(aX) = E[(aX-a\mu_X)^2] = a^2\, Var(X)$
  - $E[X+b] = \mu_X + b$

    $Var(X+b) = E[(X+b-(\mu_X+b))^2] = Var(X)$

Def. $\frac{X-\mu_X}{\sigma_X}$ is called standardized version of X:

$$E\left[\frac{X-\mu_X}{\sigma_X}\right] = 0$$

$$Var\left[\frac{X-\mu_X}{\sigma_X}\right]=1$$

(1.3) ith moment of X. $E[X^i]$

---

(2) Conditional probabilities:

Def. conditional probability of B given A:

$$P(B|A)=\begin{cases}\frac{P(AB)}{P(A)} & \text{if } P(A)\neq 0\\ \text{undefined} & \text{if } P(A)=0\end{cases}$$

. Frequency interpretation.

Suppose we ran an experiment N times. Event A happens N_A times, event $A\cap B$ happens N_{AB} times.

frequency of observing B among those experimet in which A has happened

$$= \text{frequency of observing event B give event A has happened}$$

$$=\frac{N_{AB}}{N_A}=\frac{N_{AB}}{N}\cdot\frac{N}{N_A}\approx P(AB)\cdot(P(A))^{-1}$$

. It satisfies axioms of probability: suppose $P(A)>0$

1. $P(B|A)\geq 0$
2. $P(\Omega|A)=1$
3. If $B_1, B_2, \dots$ are mutually exclusive, $P(B_1\cup B_2\cup\cdots|A)=\sum_{i=1}^{\infty}P(B_i|A)$

. Bayes' rule $P(B|A)\cdot P(A)=P(AB)=P(A|B)\cdot P(B)$

---

③ Mutually independent events.

Def. We say A and B are independent if $P(AB) = P(B)\,P(A)$

- Knowledge of A won't affect probability of B: $P(B|A) = P(B) = P(B|A^c)$