Linearity of Expectation

- Recall that $E[X] = \sum_{x \in \Omega'} x p_X(x)$.
- Functions of random variables are random variables.
- Furtermore, $E[g(X)] := \sum_{x \in \Omega'} g(x) p_X(x)$
- For $Y = aX + b$, what is $E[Y]$?

$$\begin{aligned} E[Y] &= \sum_{x \in \Omega'} (ax + b) p_X(x) \\ &= a \sum_{x \in \Omega'} x p_X(x) + b \\ &= aE[X] + b. \end{aligned}$$

- What is the PMF of Y?

PMF of Y where $Y = aX + b$.

- Suppose the range of X is $\Omega' = \{x_1, x_2, \ldots, x_n\}$. Then what is the range Ω'' of Y ?
- $\Omega'' = \{y_1, \ldots, y_n\}$ where $y_i = ax_i + b$ for $i \in \{1, 2, \ldots, n\}$.
- It is easy to see that, $p_Y(y_i) = p_X(x_i)$ for $i \in \{1, 2, \ldots, n\}$.

$$\begin{aligned} E[Y] &= \sum_{y \in \Omega''} y p_Y(y) \\ &= \sum_{x \in \Omega'} (ax + b) p_y(ax + b) \\ &= \sum_{x \in \Omega'} (ax + b) p_x(x) \\ &= aE[X] + b. \end{aligned}$$

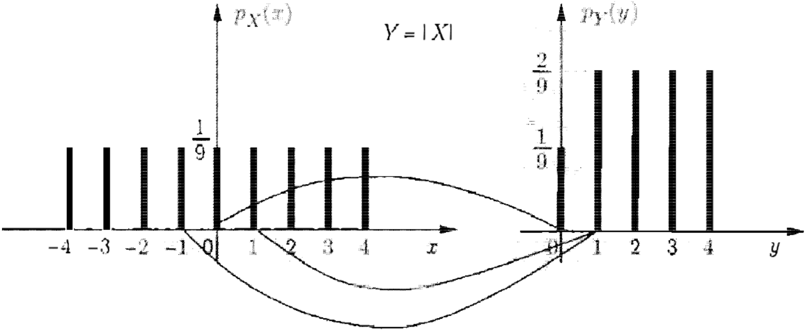
- What if $Y = g(X)$ where the function $g(.)$ is many to one? What is the PMF of Y then ?

Function of random variables

- Consider $Y = |X|$ where X is the outcome of an experiment where an integer is chosen uniformly from -4 to 4.
- $p_X(x) = \frac{1}{9}$ for $x \in \{-4, -3, \ldots, 3, 4\}$.
- What is the range Ω' for Y? $\Omega' = \{0, \ldots, 4\}$.
- What is $p_Y(2)$?
- $p_Y(2) = \sum_{\{x:|x|=2\}} p_X(x) = p_X(-2) + p_x(2) = \frac{2}{9}$.

Function of random variables

- $p_Y(2) = \sum_{\{x:|x|=2\}} p_X(x) = p_X(-2) + p_x(2) = \frac{2}{9}$.



Suppose $Y = g(X)$ and X is discrete with pmf $p_X(\cdot)$. Then $p_Y(y) = \sum_{\{x:g(x)=y\}} p_X(x)$. (Proof is HW)

$E[g(X)]$

Theorem: Suppose $Y = g(X)$ and X is discrete with pmf $p_X(\cdot)$. Then, $E[Y] = \sum_x g(x) p_X(x)$

Proof

$$\begin{aligned} E[Y] &= \sum_y y p_Y(y) \\ &= \sum_y \sum_{\{x: g(x)=y\}} g(x) p_X(x) \\ &= \sum_x g(x) p_X(x). \end{aligned}$$

$\square$

https://en.wikipedia.org/wiki/Law_of_the_unconscious_statistician

Towards Variance ..

- Recall $E[X] = \sum_{x \in \Omega'} x p_X(x)$.
- Furthermore, $E[X^n] = \sum_{x \in \Omega'} x^n p_X(x)$.
- In general, $E[g(X)] := \sum_{x \in \Omega'} g(x) p_X(x)$
- Now consider $g(X) = (X - E[X])^2$. $g(X)$ quantifies the square of the deviation of X from the mean.
- Note $g(X)$ cannot track if the deviation is positive or negative!
- $E[g(X)]$ would then tell us the mean of the square of the deviation.
- In fact, $\sqrt{E(g(X))}$ quantifies the deviation.

Variance

- $E[g(X)] = E[(X - E[X])^2]$ is called as the variance of random variable X.
- $Var(X) := E[(X - E[X])^2]$
- HW: Prove that $E[(X - E[X])^2] = E[X^2] - E[X]^2$
- $\sigma_X = \sqrt{Var(X)}$ is called as the standard deviation of X.
- For a fair coin toss, instead of $\Omega' = \{1, -1\}$, what if we use $\{+100, -100\}$? The latter has more variance!
- HW: What is $Var(Y)$ where $Y = aX + b$?

Examples of discrete random variables

Indicator random variable

- Indicator random variable $1_A(\omega) = \begin{cases} 1, & \text{If } \omega \in A \subseteq \Omega \\ 0, & \text{otherwise.} \end{cases}$
- Its PMF is $p_{1_A}(x) = \begin{cases} \mathbb{P}(A), & \text{when } x = 1 \\ 1 - \mathbb{P}(A), & \text{when } x = 0. \end{cases}$
- This is a discrete random variable even though Ω could be continuous.
- For example, Event A could be that the number picked uniformly on the real line is positive.
- What is its CDF and mean denoted by $E[1_A]$?
- What about its mean variance and moments?

Bernoulli random variable

- Bernoulli random variable $X = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{otherwise.} \end{cases}$
- This is same as an indicator variable but here we do not specify A.
- As a matter of convenience, we will start ignoring Ω from now on.
- These random variables are used in Binary classification in ML. $X = 1$ if image has a cat.
- Basic models of Multi-arm bandit problem assume Bernoulli Bandits.
- $E[X] = p, E[X^n] = p$.

Binomial $B(n, p)$ random variable.

- Consider a biased coin (head with probability p) and toss it n times.
- Denote head by 1 and tail by 0.
- Let random variable N denote the number of heads in n tosses.
- PMF of N?. $p_N(k) = \binom{n}{k} p^k (1-p)^{n-k}$.
- HW: What is $E[N], E[N^2], Var(X)$?

Geometric random variable

- Consider a biased coin (head with probability p) and suppose you keep tossing it till head appears the first time.
- Let random variable N denote the number of tosses needed for head to appear first time.
- What is the PMF of N? $p_N(k) = (1-p)^{k-1}p$.
- HW: What is $E[N], E[N^2], Var(N)$?
- What is $\bar{F}_N(k) := 1 - F_N(k) = P(N > k)$?
- What is $P(N > k + m \mid N > k)$? $(= P(N > m))$ (memoryless property)