9.1) Expectation and Variance

Vitor Kamada

February 2020

Reference

Tables, Graphics, and Figures from

Principles and Techniques of Data Science

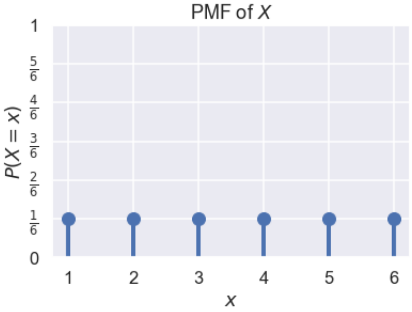
Lau et al. (2019): Ch 12 Probability and Generalization

https://www.textbook.ds100.org/ch/12/prob_exp_var.html

Probability Mass Functions (PMF)

$$1) \sum_{x \in \mathbb{X}} P(X = x) = 1$$

2) For all $x \in \mathbb{X}, 0 \leq P(X = x) \leq 1$



$$P(X = 1) = P(X = 2) = \ldots = P(X = 6) = \frac{1}{6}$$

Joint Distribution

X: # of heads in 10 coin flips

Y: # of tails in the same set of 10 coin flips

$$P(X = 0, Y = 10) = P(X = 10, Y = 0) = (0.5)^{10}$$

$$P(X = 6, Y = 6) = 0$$

Marginal Distribution

$$\sum_{y \in \mathbb{Y}} P(X = x, Y = y) = P(X = x)$$

$$\begin{aligned}
\sum_{y \in \mathbb{Y}} P(X = x, Y = y) &= \sum_{y \in \mathbb{Y}} P(X = x) \times P(Y = y \mid X = x) \\
&= P(X = x) \times \sum_{y \in \mathbb{Y}} P(Y = y \mid X = x) \\
&= P(X = x) \times 1 \\
&= P(X = x)
\end{aligned}$$

Expectation

$$\mathbb{E}[X] = \sum_{x \in \mathbb{X}} x \cdot P(X = x)$$

$$\begin{aligned}\mathbb{E}[X] &= 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + \ldots + 6 \cdot P(X = 6) \\ &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} \\ &= 3.5\end{aligned}$$

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

$$\mathbb{E}[cX] = c\mathbb{E}[X]$$

Expectation

$$\mathbb{E}[X] = \sum_{x \in \mathbb{X}} x \cdot P(X = x)$$

$$\begin{aligned}
\mathbb{E}[X] &= 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + \ldots + 6 \cdot P(X = 6) \\
&= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} \\
&= 3.5
\end{aligned}$$

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

$$\mathbb{E}[cX] = c\mathbb{E}[X]$$

Variance

X and Y are independent

$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

$$Var(aX + b) = a^2 Var(X)$$

$$Var(X + Y) = Var(X) + Var(Y)$$

Covariance

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

$$Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

If X and Y are independent, then $Cov(X, Y) = 0$

Sample Means: $\hat{p} = \frac{X_1+X_2+...+X_n}{n}$

$$
\begin{aligned}
\mathbb{E}[\hat{p}] &= \mathbb{E}\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right] \\
&= \frac{1}{n}\mathbb{E}[X_1 + \ldots + X_n] \\
&= \frac{1}{n}\left(\mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n]\right) \\
&= \frac{1}{n}(p + \ldots + p) \\
&= \frac{1}{n}(np) \\
\mathbb{E}[\hat{p}] &= p
\end{aligned}
$$

Risk or Expected Loss

$$
\begin{aligned}
R(\theta) &= \mathbb{E}[(X-\theta)^2] \\
&= \mathbb{E}\left[(X-\mathbb{E}[X]+\mathbb{E}[X]-\theta)^2\right] \\
&= \mathbb{E}\left[\left((X-\mathbb{E}[X])+(\mathbb{E}[X]-\theta)\right)^2\right] \\
&= \mathbb{E}\left[(X-\mathbb{E}[X])^2+2(X-\mathbb{E}[X])(\mathbb{E}[X]-\theta)+(\mathbb{E}[X]-\theta)^2\right]
\end{aligned}
$$

$$
= \mathbb{E}\left[(X-\mathbb{E}[X])^2\right] + 2(\mathbb{E}[X]-\theta)\underbrace{\mathbb{E}\left[(X-\mathbb{E}[X])\right]}_{=0} + (\mathbb{E}[X]-\theta)^2
$$

$$
R(\theta) = \underbrace{(\mathbb{E}[X]-\theta)^2}_{\text{bias}} + \underbrace{Var(X)}_{\text{variance}}
$$

Statistical vs Empirical Risk

$$R(\theta) = \underbrace{(\mathbb{E}[X] - \theta)^2}_{\text{bias}} + \underbrace{Var(X)}_{\text{variance}}$$

$$\mathbb{E}[X] = \sum_{x \in \mathbb{X}} x \cdot P(X = x)$$

$$\theta^* = \mathbb{E}[X]$$

$$\mathbb{E}[X] \approx \frac{1}{n} \sum_{i=1}^{n} x_i = \text{mean}(\mathbf{x})$$

$$\hat{\theta} = \text{mean}(\mathbf{x})$$