Probability

$$P(A|B) = \frac{P(A\cap B)}{P(B)} \qquad P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

If A and B are independent $P(A\cap B) = P(A)P(B)$

Expected Value

$$E[x] = \int_{-\infty}^{\infty} xp(x)\,dx \qquad E[g(x)] = \int_{-\infty}^{\infty} g(x)p(x)\,dx$$

Variance

$$Var[x] = E[(x - E[x])^2] = \sigma^2 = \int_{-\infty}^{\infty} (x - E[x])^2 p(x)\,dx$$

$$E((x - E[x])^2) = E[x^2] - (E[x])^2$$

Covariance

$$\begin{aligned} cov(X_1, X_2) &= E[(X_1 - m_1)(X_2 - m_2)] \\ &= E[(X_1)(X_2)] - m_1 m_2 \end{aligned}$$

Correlation

$$\rho = \frac{cov(X_1, X_2)}{\sqrt{V(X_1)V(X_2)}}$$

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Algebra of Random Variable

Probability Distribution

Cumulative Distribution Function (CDF)

Normal

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad E[X] = \mu \qquad Var[X] = \sigma^2$$

Exponential

$$\lambda e^{-\lambda x} \qquad E[X] = \frac{1}{\lambda} \qquad Var[X] = \frac{1}{\lambda^2}$$

Uniform

$$\begin{cases} \dfrac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = \frac{1}{2}(a + b) \qquad Var[X] = \frac{1}{12}(b - a)^2$$

Bernoulli

$$\begin{cases} q = 1 - p & \text{if } k = 0 \\ p & \text{if } k = 1 \end{cases} \qquad E[X] = p \quad Var[X] = p(1 - p)$$

Binomial

$$\binom{n}{k} p^k (1 - p)^{n-k} \qquad E[X] = np \qquad Var[X] = np(1 - p)$$

Poisson

$$\frac{\lambda^k e^{-\lambda}}{k!} \qquad E[X] = \lambda \qquad Var[X] = \lambda$$

Pareto

$$\frac{\alpha x_m^a}{x^{\alpha+1}} \qquad E[x] = \begin{cases} \infty & \text{for } \alpha \le 1 \\ \dfrac{\alpha x_m}{\alpha - 1} & \text{for } \alpha > 1 \end{cases}$$

$$Var[X] = \begin{cases} \infty & \text{for } \alpha \le 1 \\ \dfrac{x_m^2 \alpha}{(\alpha - 1)^2(\alpha - 2)} & \text{for } \alpha > 1 \end{cases}$$

MLE

To find the MLE given data

1. The likelihood function $P(data|\lambda)$, λ is the parameter
2. $\frac{d}{d\lambda}$(log likelihood) $= 0$, Find λ

Hypothesis Testing

Errors

		True State of Nature	
		H_0	H_A
Our Decision	Reject H_0	Type-I Error	correct decision
	Reject H_0	correct decision	Type-II Error

One Sample z-test

Use when the **Variance** (σ^2) of the data is known

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

One Sample t-test

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad \text{where } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad df = n - 1$$

Two Sample z-test

$$z = \frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

Two Sample t-test with Equal Variance

Assume that the data have $\sigma_1 = \sigma_2$

$$t = \frac{\bar{x} - \bar{y}}{s_p} \qquad s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n + m - 2}\left(\frac{1}{n} + \frac{1}{m}\right)$$

Two Sample t-test with Unequal Variance

Assume that the data have $\sigma_1 \neq \sigma_2$

$$t = \frac{\bar{x} - \bar{y} - \mu_0}{s_P} \qquad s_P^2 = \frac{s_x^2}{n} + \frac{s_y^2}{n}$$

$$df = \frac{(s_x^2/n + s_y^2/m)^2}{(s_x^2/n)^2/(n-1) + (s_y^2/m)^2/(m-1)}$$

Paired two-sample t-test

$$t = \frac{\bar{w} - \mu_0}{s/\sqrt{n}} \quad w_i = x_i - y_i \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(w_i - \bar{w})^2$$

A/B Testing

MDE

Sample Size

m is the split rate (50:50 = 1, 80:20 = 4)

$$n = \frac{m+1}{m}\left(\frac{(Z_\alpha + Z_\beta)\sigma}{MDE}\right)^2$$

When not to do an A/B test

- Things that cannot be summarized into one or a few metrics
- Totally new things
- Delayed results
- One-off events
- Cannot split group independently