Probability

$$0 \leq P[A] \leq 1$$

0 → impossible event

1 → certain (definite) event

- $P[A]$ the probability of event A.
- The sample space of throwing a dice is $\{1,2,3,4,5,6\}$.
- An event of throwing a dice is getting even number. (Event B)
- $P[B] = 1/2$
- $P[\bar{B}]$ is complement of event B. $P[A] + P[\bar{A}] = 1$

Q: What is the probability of getting 2 heads and 3s tail by throwing a coin?

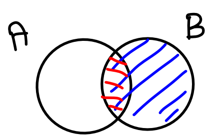
A: HHTTT → $\dfrac{5!}{2!3!} = 10$ $\qquad \dfrac{10}{2^5} = \dfrac{5}{16}$

- If event A and event B are independent events, then $P[A \cap B] = P[A] . P[B]$
- $P[A|B] = P[A]$ and $P[B|A] = P[B]$
- If A and B are independent events, then $(A, B^c), (A^c, B), (A^c, B^c)$ are independent couples, too.
- There are $2^K - K - 1$ conditions that has to apply for K number of independent events.

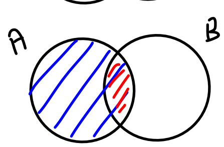
Ex: If A, B, C are independent events, then $P[A \cap B] = P[A] . P[B]$, $P[A \cap C] = P[A] . P[C]$, $P[B \cap C] = P[B] . P[C]$

$P[A \cap B \cap C] = P[A] \cdot P[B] \cdot P[C]$

Conditional Probability



$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$ (The probability of A when the event B is given.)



$$P[B|A] = \frac{P[A \cap B]}{P[A]}$$

Disjoint Events: A and B are disjoint events. $P[A \cap B] = \emptyset$ If $A = \overline{B}$ then $P[A \cup B] = 1$

Bayes Rule: $P[B_i | A] = \dfrac{P[A | B_i] \cdot P[B_i]}{P[A]} \Rightarrow P[B_i | A] = \dfrac{P[A | B_i] \cdot P[B_i]}{\sum_{i=1}^{m} P[A | B_i] \cdot P[B_i]}$

- $P[A \cap B \cap C] = P[C | A \cap B] \cdot P[A \cap B] = P[C | A \cap B] \cdot P[A | B] \cdot P[B]$

- If $A \subset B$, then $P[B | A] = 1$.

Law of Total Probability: $P[A] = \sum_{i=1}^{m} P[A | B_i] \cdot P[B_i]$

Chapter 2:

Probability Mass Function: Discrete Probability Distribution

For a discrete random variable x with PMF $P_x(x)$ and range S_x,

- $\sum_{x \in S_x} P_x(x) = 1$

- $P_x(x) \geq 0$

Geometric Random Variable: $P_x(x) = \begin{cases} p \cdot (1-p)^{x-1} & x = 1, 2, 3, \ldots \\ 0 & \text{otherwise.} \end{cases}$ $E[X] = 1/p$ $\sigma^2 = \dfrac{1-p}{p^2}$ $F(k) = P(x \leq k) = 1 - q^k$, $q = 1-p$

Binomial Random Variable: $P_x(x) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$, $n \geq 1$ $E[X] = n \cdot p$ $\sigma^2 = n p \cdot (1-p)$

Pascal Random Variable: $P_x(x) = \binom{n-1}{k-1} \cdot p^{k-1} \cdot (1-p)^{n-k} \cdot p$ $E[x] = \dfrac{n-1}{p}$

Expected Value ($E[X]$): $\mu = \sum_{x \in S_x} x \cdot P_x(x)$

Variance ($Var[X]$): $\sigma^2 = E[x^2] - (E[x])^2 \rightarrow E[x^2] = \sum_{x \in S_x} x^2 \cdot P_x(x)$

Standard Deviation (σ): $\sigma = \sqrt{Var[X]}$

Probability Density Function : Continious Probability Distribution

- $f_X(x) \geq 0$
- $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
- $E[X] = \mu = \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx$
- $Var[X] = \sigma^2 = E[X^2] - (E[X])^2 \longrightarrow E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f_X(x)\,dx$

Properties of E[X] and Var[X]

E[X]

- $E[a] = a$ a is constant value.
- $E[kX] = k \cdot E[X]$
- $E[mX+n] = m \cdot E[X] + n$

Var[X]

- $Var[a] = 0$
- $Var[kX] = k^2 Var[X]$
- $Var[mX+n] = m^2 Var[X]$

Cumulative Distribution Function (CDF)

Moment Generating Function

$$M_X(t) = \sum_{x \in S_X} e^{xt} \cdot P_X(x) = E[e^{xt}] \longrightarrow PMF$$

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x)\,dx = E[e^{xt}]$$

$$\frac{d}{dt}(M_X(t))\Big|_{t=0} = E[X]$$

$$\vdots$$

$$\frac{d^n}{dt^n}(M_X(t))\Big|_{t=0} = E[X^n]$$

Conditional Probability Mass Function

$$P_X(x) = \begin{cases} 0.2 & x = 1, \\ 0.3 & x = 2, \\ 0.5 & x = 3, \\ 0 & \text{otherwise.} \end{cases}$$

$A = \{X > 1\}$ is given.

$$P_{X|A}(x) = \begin{cases} P_X(x) / P(A) & x = 2, 3, \\ 0 & \text{otherwise.} \end{cases}$$

$$E_{X|A} = \sum_{x \in S_X} x_k \cdot P_{X|A}(x_k)$$

$$\sigma^2 = E[X^2|A] - E[X|A]^2$$

$$P_{X|A}(x) = \begin{cases} 0.375 & x = 2, \\ 0.625 & x = 3, \\ 0 & \text{otherwise.} \end{cases}$$

Chebyshev Inequality

- $P(|X+\mu| \geqslant k.\sigma) \leqslant \frac{1}{k^2}$

- $P(|X+\mu| < k.\sigma) \geqslant 1-\frac{1}{k^2}$

Poisson Random Variable

$P_X(k) = \frac{\lambda^k}{k!} . e^{-\lambda}$ $k=0,1,\ldots$ $\lambda = n.p$ (We use this to calculate hard binomial problems.)

$P_X(k) = \frac{a^k}{k!} . e^{-a}$ $k=0,1,\ldots$ $a = \lambda . T$ (If there is a time constraint we use this.)

- $E[X] = \lambda$
- $Var[X] = \lambda$

- Moment Function of Poisson: $e^{\lambda(e^t-1)}$

Uniform Distribution

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leqslant x \leqslant b, \\ 0 & \text{otherwise.} \end{cases}$$

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

Exponential Distribution

$$f(x) = \begin{cases} \lambda . e^{-\lambda . x} & x \geqslant 0, \\ 0 & x < 0 \end{cases} \qquad F(x) = \begin{cases} 1 - e^{-\lambda . x} & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

- $E[X] = \frac{1}{\lambda}$ $\sigma^2 = \frac{1}{\lambda^2}$

Statistic

confidence level: $(1-\alpha)$ Aralık tahminin popülasyon parametresini içermesi konusundaki kesinlik düzeyi

confidence interval: Güven düzeyine göre aralık tahmini.

significance level: (α) Gerçek değerin tahmin aralığı dışında olma olasılığı.

error: Güven aralığının kapsadığı nokta tahminden muhtemel en büyük uzaklıktır. $\bar{x}-e < \mu < \bar{x}+e$

The Confidance Interval when the Population Variant is Known

$$\bar{x} - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$\bar{x}$ = mean of sample

σ = standard deviation of population (not sample)

n = number of sample data

α = significance level

μ = mean of population

$$Z_t = \frac{x-\mu}{\frac{\sigma}{\sqrt{n}}}$$

- To find $Z_{\frac{\alpha}{2}}$ → find $1-\frac{\alpha}{2}$, then find corresponding value on Z-table

The Confidence Interval When the Population Variant is Unknown $(n \geq 30)$

$$\bar{x} - Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

s = standard deviation of sample

$$Z_t = \frac{x-\mu}{\frac{s}{\sqrt{n}}}$$

The Confidance Interval When the Population Variant is Unknown $(n < 30)$

$$\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

$$t_t = \frac{x-\mu}{\frac{s}{\sqrt{n}}}$$

- To find $t_{\frac{\alpha}{2}}$ → find $\frac{\alpha}{2}$, find degrees of freedom (DF) → $v = n-1$, then $t_{\frac{\alpha}{2},(v)}$

- For only bottom interval or top interval → Z_α instead of $Z_{\frac{\alpha}{2}}$, t_α instead of $t_{\frac{\alpha}{2}}$

- $e = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

- $n = \left(\frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{e}\right)^2$ or $n = \left(\frac{t_{\frac{\alpha}{2}} \cdot \sigma}{e}\right)^2$

- $\sigma = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$

Confidence Interval for Variance

$$\frac{(n-1)\cdot s^2}{\chi^2_{\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)\cdot s^2}{\chi^2_{1-\frac{\alpha}{2}}} \quad , \quad \sqrt{\frac{(n-1)\cdot s^2}{\chi^2_{\frac{\alpha}{2}}}} < \sigma < \sqrt{\frac{(n-1)\cdot s^2}{\chi^2_{1-\frac{\alpha}{2}}}}$$

- Chi-Square table read exactly like t-table. ($v = n-1$, freedom)

$$\chi^2_t = \frac{n-1 \cdot s^2}{\sigma^2}$$

Confidence Interval for Proportion

$$\hat{p} - Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} < p < \hat{p} + Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

p: popülasyon oranı

$\hat{p}$: örneklem oranı

$$Z_t = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Confidence Interval for Paired Observations

$$\bar{d} - t_{\frac{\alpha}{2}} \cdot \frac{S_d}{\sqrt{n}} < \mu_d < \bar{d} + t_{\frac{\alpha}{2}} \cdot \frac{S_d}{\sqrt{n}}$$

Confidence Interval for 2 Population (Population Variance is Known)

$$(\bar{x}_1 - \bar{x}_2) - Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Confidence Interval for 2 Population (Population Variances are Unknown and not Equal)

$$(\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2}, v} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2}, v} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Confidence Interval for 2 Population (Population Variances are Unknown but Equal)

$$(\bar{x}_1-\bar{x}_2) - t_{\frac{\alpha}{2},v} \cdot s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1-\mu_2 < (\bar{x}_1-\bar{x}_2) + t_{\frac{\alpha}{2},v} \cdot s_p \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$v = n_1+n_2-2$$

$$s_p^2 = \frac{(n_1-1)\cdot s_1^2+(n_2-1)\cdot s_2^2}{n_1+n_2-2}$$

Confidence Interval for Difference in Proportions

$$(\hat{p}_1-\hat{p}_2) - Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} < p_1-p_2 < (\hat{p}_1-\hat{p}_2) + Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Confidence Interval for Ratio of Variances

$$\frac{1}{F_{\frac{\alpha}{2},(n_1-1,n_2-1)}} \cdot \frac{s_1^2}{s_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < F_{\frac{\alpha}{2},(n_2-1,n_1-1)} \cdot \frac{s_1^2}{s_2^2}$$

Least Squares Method for Linear Regression

$$\hat{y} = \beta_0+\beta_1 x \qquad \beta_0 = \bar{y}-\beta_1\bar{x} \qquad \beta_1 = \frac{S_{xy}}{S_{xx}} \qquad S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} \qquad S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n}$$

$$\sigma^2 = \frac{1}{n-2}\cdot\sum_{i=1}^{n} e_i^2$$

$$\downarrow$$

$$\sigma^2 = \frac{1}{n-2}\cdot\sum_{i=1}^{n} (y_i-\hat{y}_i)^2$$

y_i: given y values

$\hat{y}_i$: calculated y values acc. to found equation

Göreceli Olasılık

$$O(A) = \frac{P(A)}{P(A^c)}$$

- $0 \leq O(A) < \infty$
- $O(A^c) = \dfrac{1}{O(A)}$
- $P(A) = \dfrac{O(A)}{1+O(A)}$

Joint Probability

- $a \in X$, $b \in Y$ every a and b, if $P(X=a, Y=b) = P(X=a)\cdot P(Y=b)$, then X and Y are independent random variables.
- Corelation → $r_{xy} = E[XY] = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i \cdot y_j \cdot P(X=x_i, Y=y_j)$
- Covariance → $\sigma_{xy} = Cov(X,Y) = E[XY]-E[X]\cdot E[Y]$
- Corelation constant → $\rho_{xy} = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$

Confidence Interval Pop. Var. is Known Z

Confidence Interval Pop. Var is Unknown ($n \geq 30$) Z

Confidence Interval Pop. Var is Unknown ($n < 30$) t

Confidence Interval for proportion Z

Confidence Interval for Variance χ

Confidence Interval for Paired t

Confidence difference When Pop. var. is known Z

Confidence difference when pop. var is unknown and unequal t

Confidence difference when pop. var. is " " equal t

Proportion difference Z

$$(\bar{X}_1 - \bar{X}_2) - t_{\frac{\alpha}{2}, v} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{\frac{\alpha}{2}, v} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$