Computational Statistics Formulae

Random Variable

- X $\to$ function:
  input: $\Omega : \{\xi \in \Omega\}$ where ξ one possible event.
  output: $R_X \to$ (Support of X)

Discrete	Continuous
PMF: $P_x(x) = P[X = x]$ **CDF:** $F_X(x) = P[X \leq x]$	**PMF:** $f(x)$ **CDF:** $F(x) = \int_{-\infty}^{x} f(t)dt$
Expectation $\mathbb{E}[X] = \sum x \cdot P_X(x) = \mu$ **Variance** $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[(X-\mu)^2]$ $\mathbb{E}[X^2] = \sum x^2 \cdot P_X(x)$ $\mathbb{E}[X]$ is **linear**. $\mathbb{E}[c] = c$ $\text{Var}[X + c] = \text{Var}[X]$	**Expectation** $\mathbb{E}[X] = \int_{R_X} x \cdot f(x)dx$ **Variance** $\text{Var}[X] = \int_{R_X} d^2 f(x)dx$ where, $d = x - \mu$

Discrete Probability Distribution

Binomial Distribution

Notation: $X \sim \text{Bin}(n, p)$

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^{n-r} b^r$$

$$P_X(r) = \binom{n}{r} q^{n-r} p^r$$

$E[X] = np$
$\text{Var}[X] = npq$

Bernoulli Distribution

$X \sim \text{Bin}(1, p)$ is a Bernoulli Distribution.

Poisson Distribution

Notation: $X \sim \text{Poisson}(\lambda)$

$$P_x(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$\mathbb{E}[X] = \lambda = \text{Var}[X]$$

λ: avg. rate of event occuring
x: no. of occuring events

Example:
On an avg. if 20 people visit a museum/hr, what is the probability that x no. of people visit the museum in one hour?

Geometric Distribution

Notation: $X \sim \text{Geom}(p)$

2 def:
1. X: # failures before 1^{st} success
2. Y: # failures until 1^{st} success

Clearly, $Y = X + 1$

Note

$$\sum_{0}^{\infty} kp^k = \frac{p}{q^2} \qquad \sum_{0}^{\infty} k^2 p^{k-1} = \frac{p(1+p)}{q^2}$$

- $P[X = x] = q^x p$
- $P[Y = y] = q^{y-1} p$
- $F_{X(x)} = 1 - q^{x+1}$
- $F_{Y(y)} = 1 - q^y$
- $\mathbb{E}[X] = q/p$
- $\mathbb{E}[Y] = 1/p$
- $\text{Var}[X \text{ or } Y] = q/p^2$

Negative Binomial Distribution

Notation: $X \sim \text{Geom}(p)$

2 def:

1. X: # failures before r^{th} success
2. Y: # failures until r^{th} success

Clearly, $Y = X + r$

Note,
-ve Binomial Distribution is the sum of r Geometric Distribution. i.e.:

$$X = X_1 + X_2 + X_3 + \dots + X_r$$

$$Y = Y_1 + Y_2 + Y_3 + \dots + Y_r$$

$$P[X = x] = \binom{x+r-1}{r-1} q^x p$$

$$P[Y = y] = \binom{y-1}{r-1} q^{y-1} p$$

$F_{X(x)} = 1 - q^{x+1}$
$F_{Y(y)} = 1 - q^y$

$\mathbb{E}[X] = rq/p$
$\mathbb{E}[Y] = r/p$
$\text{Var}[X \text{ or } Y] = rq/p^2$

Hypergeometric Distribution

Notation: $X \sim \text{HG}(N, n, m)$
where,
$N \to$ Total population
$n \to$ Sample Size
$m \to$ Total success in N
$x \to$ Total success in m

$$P_X(x) = \frac{\binom{n}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$$

Continuos Probability Distributions

$\lambda \to$ avg. **rate parameter**, $\beta \to$ **scale parameter** (avg. waiting time), $\alpha \to$ **shape parameter** (# required events).

Uniform Distribution

Notation: $X \sim \text{Uniform}(a, b)$

$$f(x) = \begin{cases} f(x) & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

$$\mu = \frac{a+b}{2}, \quad \sigma^2 = \frac{(b-a)^2}{12}$$

$$F_X(x) = \frac{x-a}{b-a}$$

$$x_m = \frac{a+b}{2}$$

Exponential Distribution

Notation: $X \sim \text{Exponential}(\lambda)$

Holds Memoryless property

$X \to$ waiting time before another event

$$f(x) = \lambda e^{-\lambda x}$$

$$\mathbb{E}[X] = \frac{1}{\lambda}, \quad \text{Var}[X] = \frac{1}{\lambda^2}$$

$$\beta = \frac{1}{\lambda}$$

Gamma Distribution

Notation: $X \sim \text{Gamma}(p)$

Sum of α exponential distribution and events occur independently over time.

Holds Memoryless property

$X \to \alpha$ events in x time if avg rate is λ.
$\beta \to$ avg. waiting time for next event.

$$\Gamma(\alpha) = (\alpha-1)! \text{ or } \int_0^\infty t^{\alpha-1} e^{-t} dt$$

$$\gamma(\alpha, \lambda x) = \int_0^{\lambda x} t^{\alpha-1} e^{-t} dt$$

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$$

$$\mathbb{E}[X] = \alpha\beta, \quad \text{Var}[X] = \alpha\beta^2$$

$$F_X(x) = \frac{\gamma(\alpha, \lambda x)}{\Gamma(\alpha)}$$

We get Exponential Distribution if $\alpha = 1$

We get Erlang Distribution if $\alpha \in \mathbb{Z}_+$ where,

$$F_X(x) = 1 - \sum_{n=0}^{k-1} \frac{1}{n!} e^{-\lambda x} (\lambda x)^n$$

Normal Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

$$Z = \frac{X-\mu}{\sigma} \text{ and } \mathbb{E}[Z] = 0, \text{Var}[Z] = 1$$

Normal Approximation

Mind the correction factor.

$$P[a \le X \le b] = P[a - 0.5 \le X \le b + 0.5]$$
$$P[a < X < b] = P[a + 0.5 < X < b - 0.5]$$

Binomial Distribution

When n is large and p or q is small, We approximate probability with $\mu = np$ and $\sigma = \sqrt{npq}$.

Poisson Distribution

When λ is large, We approximate probability with $\mu = \sigma^2 = \lambda$

Log-Normal Distribution

X is log-normal $\to Y = \ln(X)$ is normal $\to X = \exp(Y)$ is log-normal
$X \sim \text{Log-Normal}(\mu, \sigma^2)$ $\mathbb{E}[Y] = \mu, \text{Var}[Y] = \sigma^2$.

$\mathbb{E}[X] = e^{\mu + \frac{\sigma^2}{2}} \quad \text{Var}[X] = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$

$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}, x \in R_+$

Weibull Distribution

$$f(x) = \alpha\lambda^\alpha x^{\alpha-1} e^{-(\lambda x)^\alpha}$$

$$F(x) = 1 - e^{-(\lambda x)^\alpha}$$

$$\mu = \beta \cdot \Gamma\left(1 + \frac{1}{\alpha}\right)$$

$$\sigma^2 = \beta^2 \left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma\left(1 + \frac{1}{\alpha}\right)^2\right]$$

Beta Distribution

Notation: $X \sim \text{Beta}(\alpha, \beta)$

Used in Bayesian statistics and modeling probabilities.

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < x < 1$$

where, $B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

$$\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}, \quad \text{Var}[X] = \alpha\frac{\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Cauchy Distribution

Notation: $X \sim \text{Cauchy}(x_0, \gamma)$

Heavy-tailed distribution with undefined mean and variance.

$$f(x) = \frac{1}{\pi\gamma\left(1 + \left(\frac{x - x_0}{\gamma}\right)^2\right)}$$

$$F(x) = \frac{1}{\pi}\arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}$$

Note:

- $\mathbb{E}[X]$ is **undefined**
- $\text{Var}[X]$ is **undefined**

Find the level curve at z=30 for the function:

$$f(x, y) = 100 - x^2 - y^2 = c$$