Probability for Data Science - Formula Sheet 2

tbimbato

Uniform Distribution

$X \sim Unif(a,b)$

- **Discrete:** $P(X=x) = \frac{1}{b-a+1}, \quad x \in \{a,\dots,b\}$
- **Continuous:** $f_X(x) = \frac{1}{b-a}, \quad a \le x \le b$
  otherwise $f_X(x) = 0$
- **CDF :** $F_X(x) = P(X \le x)$**:**
  - Discrete: $\frac{\lfloor x \rfloor - a+1}{b-a+1}, \quad a \le x \le b$
  - Continuous: $\frac{x-a}{b-a}, \quad a \le x \le b$
- $\mathbb{E}[X] = \frac{a+b}{2}, \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$
- **Linear Transf:** $X \sim \mathrm{Unif}(a,b) \to zX + j \sim \mathrm{Unif}((az+j),(bz+j))$
- **Sum:** $X_1, X_2 \sim \mathrm{Unif}(a,b) \Rightarrow X_1 + X_2 \sim \mathrm{Tri}(2a, 2b, a+b)$

Binomial Distribution (Number of successes in n trials)

$X \sim Bin(n,p)$

- **Bernoulli:** Special case in binomial where $n = 1$
  - $P(X=x) = \quad p$ if $x = 1, \quad (1-p)$ if $x = 0$
  - $\mathbb{E}[X] = p, \quad \mathrm{Var}[X] = p(1-p)$
  - Sum of n i.i.d. Bernoulli: Binomial(n,p)
- PMF: $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$
- $\mathbb{E}[X] = n \cdot p, \quad \mathrm{Var}[X] = n \cdot p \cdot (1-p)$
- $\mathrm{Bin}(n,p) \approx N(np, np(1-p))$ for large n

Exponential Distribution (Time Until First Event)

$X \sim Exp(\lambda)$

- **PDF:** $f(x) = \lambda e^{-\lambda x}, \quad x \ge 0$
- **CDF:** $P(X \le x) = 1 - e^{-\lambda x}$
- **Survival func:** $P(X > x) = e^{-\lambda x}$
- **Moments:**
  $\mathbb{E}[X] = \frac{1}{\lambda}, \quad \mathrm{Var}[X] = \frac{1}{\lambda^2}, \quad \mathbb{E}[X^2] = \frac{2}{\lambda^2}$
- **Memoryless prop.**: $P(X > t+s \mid X > s) = P(X > t)$
- **If** $X \sim Exp(\lambda)$ **and** $Y \sim Exp(\mu)$:
  - $\min(X,Y) \sim Exp(\lambda+\mu)$
  - $P(X<Y) = \frac{\lambda}{\lambda+\mu}, \quad P(X>Y) = \frac{\mu}{\lambda+\mu}$
  - $J = X+Y \sim Gamma(2, \lambda+\mu)$
  - If summing n i.i.d. $\mathrm{Exp}(\lambda)$ RVs, then $J \sim Gamma(n,\lambda)$ (same rate λ)
- **Scaling property:** $aX \sim Exp\left(\frac{\lambda}{a}\right)$ for $a > 0$

Poisson Distribution (How many events in a time frame?)

$X \sim Pois(\lambda)$

- **PMF:** $P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots$
- $\mathbb{E}[X] = \lambda, \quad \mathrm{Var}[X] = \lambda$
- **If** $X \sim Pois(\lambda)$ **and** $Y \sim Pois(\mu)$:
  - $X + Y \sim Pois(\lambda+\mu)$
  - Interval between events is $Exp(\lambda)$
  - $P(X+Y>k)$ If $\lambda, \mu > 10$, use normal approximation: $\mathcal{N}(\lambda+\mu, \lambda+\mu)$, where mean $= \lambda+\mu$, variance $= \lambda+\mu$

Gamma RV (Time until n events in a Poisson Process)

$X \sim \Gamma(\alpha, \lambda)$

- **PDF:** $f(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad x > 0$
- **Mean and Variance:**
  $\mathbb{E}[X] = \frac{\alpha}{\lambda}, \quad \mathrm{Var}[X] = \frac{\alpha}{\lambda^2}$
- **Special case:** If $X \sim \Gamma(1,\lambda)$, then $X \sim \mathrm{Exp}(\lambda)$
  $X \sim \Gamma(\alpha = nofevent \quad, \lambda = rate)$

Geometric RV (Trials until first success)

Indipendent bernoulli trials

$X \sim Geom(p)$

- PMF: $P(X=k) = (1-p)^{k-1} p, \quad k = 1, 2, \dots$
- $\mathbb{E}[X] = \frac{1}{p}$, $\mathrm{Var}[X] = \frac{1-p}{p^2}$
- $P(X > k) = (1-p)^k$
- $P(X \le k) = 1 - (1-p)^k$
- Memoryless prop: $P(X > k+n \mid X > k) = P(X > n)$

Normal Distribution

- **PDF:**
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
- **Mean and Variance:**
$$\mathbb{E}[X] = \mu, \quad \mathrm{Var}[X] = \sigma^2$$
- **Standardization:** $Z = \frac{X-\mu}{\sigma}$
- **Probability Computation:** with $\Phi(z) = P(Z \le z)$
$$P(X \le a) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$
$$P(X > a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$$
$$\star \quad P(a \le X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$
- **Linear t.:** $aX + b \sim N((a\mu + b), (a^2\sigma^2))$
- **Simmetry:** $P(|X| > a) = 2P(X > a)$
- **Continuity C. Discrete → normal Approximations:**
  - $P(X \le k) \approx P(X < k + 0.5) = \Phi\left(\frac{k+0.5-\mu}{\sigma}\right)$
  - $P(X \ge k) \approx P(X > k - 0.5) = 1 - \Phi\left(\frac{k-0.5-\mu}{\sigma}\right)$
  - $P(a \le X \le b) \approx P((a-0.5) \le X \le (b+0.5)) = \star$
- Sum of 2 Normally distributed RVs:
  $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$:
  $X \pm Y \sim N(\mu_1 \pm \mu_2, \sigma_1^2 \pm \sigma_2^2)$
  sum of VARIANCE not sum of ST Dev.

Sums and Transformations of RVs

- If X and Y are independent, the PDF of $J = X+Y$ is: $f_J(j) = \int_{-\infty}^{\infty} f_X(x) f_Y(j-x)\,dx$