# Probability & Statistics review

Yuanning Li

BME 2111
Neural Signal Processing and Data Analysis
2023 Fall

# Random variables

- A random variable is a map $X$ from a set $\Omega$ (equipped with a probability $P$) to $\mathbb{R}$.
- We write $P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\})$
- We write $X \sim P$ to mean that $X$ has distribution $P$.
- Suppose that $X \sim P$ and $Y \sim Q$. We say that $X$ and $Y$ have the same distribution if $P(X \in A) = Q(Y \in A)$ for all $A$. In that case we say that $X$ and $Y$ are *equal in distribution* and we write $X \stackrel{d}{=} Y$.

# CDF, PMF and PDF

- The *cumulative distribution function* (CDF) of $X$ is
  - $F_X(x) = F(x) = P(X \leq x)$
- If $X$ is discrete, its *probability mass function* (PMF) is
  - $p_X(x) = p(x) = P(X = x)$
- If $X$ is continuous, then its *probability density function* (PDF) satisfies

$$P(X \in A) = \int_A p_X(x)dx = \int_A p(x)dx \quad \text{and} \quad p_X(x) = p(x) = F'(x)$$

- The following are all equivalent:
  - $X \sim P,\ X \sim F,\ X \sim p$

# Independence

- $X$ and $Y$ are *independent* if and only if $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$ for all $A$ and $B$
- Theorem: Let $(X, Y)$ be a bivariate random vector with $p_{X,Y}(x, y)$. $X$ and $Y$ are independent if and only if $p_{X,Y}(x, y) = p_X(x)p_Y(y)$.
- $X_1,\ldots,X_n$ are independent if and only if $\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i)$
- Thus $p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_{X_i}(x_i)$
- If $X_1, \ldots, X_n$ are independent and identically distributed we say they are iid (or that they are a random sample) and we write $X_1, \ldots, X_n \sim P$ or $X_1, \ldots, X_n \sim F$ or $X_1, \ldots, X_n \sim P$

# Expected Values

- The mean or expected value of $g(X)$ is
- $\mathbb{E}(g(X)) = \int g(x)dF(x) = \int g(x)dP(x) = \begin{cases} \int_{-\infty}^{\infty} g(x)p(x)dx & \text{if } X \text{ is continuous} \\ \sum_j g(x_j)p(x_j) & \text{if } X \text{ is discrete} \end{cases}$
- Expectation is a linear operator:
  - $\mathbb{E}(\sum_{j=1}^{k} c_j g_j(X)) = \sum_{j=1}^{k} c_j \mathbb{E}(g_j(X))$
  - If $X_1, \ldots, X_n$ are independent then $\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}(X_i)$

# Expected Values

- We often write $\mu = \mathbb{E}(X)$
- $\sigma^2 = \text{Var}(X) = \mathbb{E}((X - \mu)^2)$ is the **Variance**. $\text{Var}(X) = \mathbb{E}(X^2) - \mu^2$
- If $X_1, \ldots, X_n$ are independent then $\text{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i)$
- The covariance is $\text{Cov}(X, Y) = \mathbb{E}((X - \mu_x)(Y - \mu_y)) = \mathbb{E}(XY) - \mu_x \mu_y$ , and the correlation is $\rho(X, Y) = \text{Cov}(X, Y)/(\sigma_x \sigma_y)$, $-1 \leq \rho(X, Y) \leq 1$

# Expected Values

- The *conditional expectation* of $Y$ given $X$ is the random variable $\mathbb{E}(Y|X)$ whose value, when $X = x$ is $\mathbb{E}(Y|X = x) = \int yp(y|x)dy$, where $p(y|x) = p(x, y)/p(x)$.
- The *Law of Total Expectation* or *Law of Iterated Expectation*:
  - $\mathbb{E}(Y) = \mathbb{E}[\mathbb{E}(Y|X)] = \int \mathbb{E}(Y|X = x)p_X(x)dx$
- The Law of Total Variance is:
  - $\text{Var}(Y) = \text{Var}[\mathbb{E}(Y|X)] + \mathbb{E}[\text{Var}(Y|X)]$

# The moment generating function (MGF)

- The moment generating function (MGF) is $M_X(t) = \mathbb{E}(e^{tX})$
- If $M_X(t) = M_Y(t)$ for all $t$ in an interval around 0, then $X \stackrel{d}{=} Y$.
- Check that $M_X^{(n)}(t)|_{t=0} = \mathbb{E}(X^n)$

# Transformations

- Let $Y = g(X)$. Then $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \int_{A(y)} p_X(x)dx$, where $A_y = \{x : g(x) \leq y\}$.
  Then $p_Y(y) = F'_Y(y)$.
- If $g$ is monotonic, then $p_Y(y) = p_X(h(y))|\frac{dh(y)}{dy}|$, where $h = g^{-1}$
- Example: Let $p_X(x) = e^{-x}$ for $x > 0$. Hence $F_X(x) = 1 - e^{-x}$. Let $Y = g(X) = \log X$. Then

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) = P(\log(X) \leq y) \\
&= P(X \leq e^y) = F_X(e^y) = 1 - e^{-e^y}
\end{aligned}
$$

  And $p_Y(y) = e^y e^{-e^{-y}} = e^{1-e^{-y}}$ for $y \in \mathbb{R}$

# Transformations

- Let $Z = g(X, Y)$. For example $Z = X + Y$ or $Z = X/Y$. Then we find the PDF of $Z$ as follows:
  - For each $z$, find the set $A_z = \{(x, y) : g(x, y) \leq z\}$
  - Find the CDF:
    $$F_Z(z) = P(Z \leq z) = P(g(X, Y) \leq z)$$
    $$= P(\{(x, y) : g(x, y) \leq z\}) = \int\int_{A_z} p_{X,Y}(x, y)dxdy$$
  - The PDF is $p_Z(z) = F_Z'(z)$

# Important Distributions

- Normal (Gaussian) distribution:
  - $X \sim \mathcal{N}(\mu, \sigma^2)$ if $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
  - (Multivariate) If $X \in \mathbb{R}^d$ then $X \sim \mathcal{N}(\mu, \Sigma)$ if

$$p(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

# Important Distributions

- Bernoulli distribution:
  - $X \sim$ Bernoulli$(\theta)$ if $\mathbb{P}(X = 1) = \theta$ and $\mathbb{P}(X = 0) = 1 - \theta$ and hence $p(x) = \theta^x(1 - \theta)^{1-x}, x = 0,1$
- Binomial distribution:
  - $X \sim$ Binomial$(\theta)$ if $p(x) = \mathbb{P}(X = x) = \binom{n}{x} \theta^x(1 - \theta)^{n-x}$, $x \in \{0,\ldots,n\}$
- Uniform distribution:
  - $X \sim$ Uniform$(0,\theta)$ if $p(x) = \mathbf{I}(0 \leq x \leq \theta)/\theta$

# Sample Mean and Variance

- The sample mean is $\overline{X} = \dfrac{1}{n} \sum_i X_i$
- The sample variance is $S^2 = \dfrac{1}{n-1} \sum_i (X_i - \overline{X})^2$
- Let $X_1, \ldots, X_n$ be i.i.d. with $\mu = \mathbb{E}(X_i) = \mu$ and $\sigma^2 =$ Var$(X_i) = \sigma^2$.

  Then $\mathbb{E}(\overline{X}) = \mu$, Var$(\overline{X}) = \dfrac{\sigma^2}{n}$, $\mathbb{E}(S^2) = \sigma^2$.

# Sample Mean and Variance

- If $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, then
  - $\overline{X} = \sim \mathcal{N}(\mu, \dfrac{\sigma^2}{n})$
  - $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$
  - $\overline{X}$ and $S^2$ are independent

# Delta Method

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, $Y = g(X)$ and $\sigma^2$ is small, then
    - $Y \approx \mathcal{N}(g(\mu), \sigma^2(g'(\mu))^2)$
- To see this, note that for some $\xi$
    - $Y = g(X) = g(\mu) + (X - \mu)g'(\mu) + \dfrac{(X-\mu)^2}{2} g''(\xi)$
- Now $\mathbb{E}((X-\mu)^2) = \sigma^2$ which we are assuming is small and so
    - $Y = g(X) \approx g(\mu) + (X - \mu)g'(\mu)$
- Thus
    - $\mathbb{E}(Y) \approx g(\mu)$, Var$(Y) \approx (g'(\mu))^2\sigma^2$
- Hence
    - $Y = g(X) \approx \mathcal{N}(g(\mu), \sigma^2(g'(\mu))^2)$

# Useful Facts

- $\displaystyle\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$
- $\displaystyle\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$
- Geometric series: $a + ar + ar^2 + \ldots = \dfrac{a}{1-r}$, for $0 < r < 1$
- Partial Geometric series $a + ar + ar^2 + \ldots + ar^{n-1} = \dfrac{a(1-r^n)}{1-r}$
- Binomial Theorem: $\displaystyle\sum_{x=0}^{n} \binom{n}{x} a^x = (1+a)^n, \quad \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x} = (a+b)^n$
- Hypergeometric identity: $\displaystyle\sum_{x=0}^{\infty} \binom{a}{x} \binom{b}{n-x} = \binom{a+b}{n}$

# Common Distributions: Uniform

- $X \sim U(1,\ldots,N)$
- $X$ takes values $x = 1,2,\ldots,N$
- $P(X = x) = \dfrac{1}{N}$
- $\mathbb{E}(X) = \sum_x xP(X = x) = \sum_x x\dfrac{1}{N} = \dfrac{N+1}{2}$
- $\mathbb{E}(X^2) = \sum_x x^2P(X = x) = \sum_x x^2\dfrac{1}{N} = \dfrac{(N+1)(2N+1)}{6}$
- $Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \dfrac{N^2-1}{12}$

# Common Distributions: Binomial

- $X \sim \text{Bin}(n, p)$
- $X$ takes values $x = 0,1,\ldots,n$
- $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$
- $\mathbb{E}(X) = np$
- $Var(x) = np(1 - p)$

# Common Distributions: Geometric

- $X \sim \text{Geom}(p)$
- $P(X = x) = p(1-p)^{x-1}, x = 1,2,\ldots$
- $\mathbb{E}(X) = \sum_x x(1-p)^{x-1} = p \sum_x \frac{d}{dp}(-(1-p)^x) = p\frac{d}{dp}(-\frac{1}{p}) = \frac{1}{p}$

# Common distribution: Poisson

- $X \sim$ Poisson$(\lambda)$, with parameter $\lambda > 0$
- $P(X = k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$, where $k$ is the number of occurrences $(k = 0,1,2,\ldots)$
- Mean: $\mathbb{E}[X] = \lambda$

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} kP(X = k) = \sum_{k=0}^{\infty} k \frac{1}{k!} \lambda^k e^{-\lambda}$$

$$= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{1}{(k-1)!} \lambda^{k-1} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

# Common distribution: Poisson

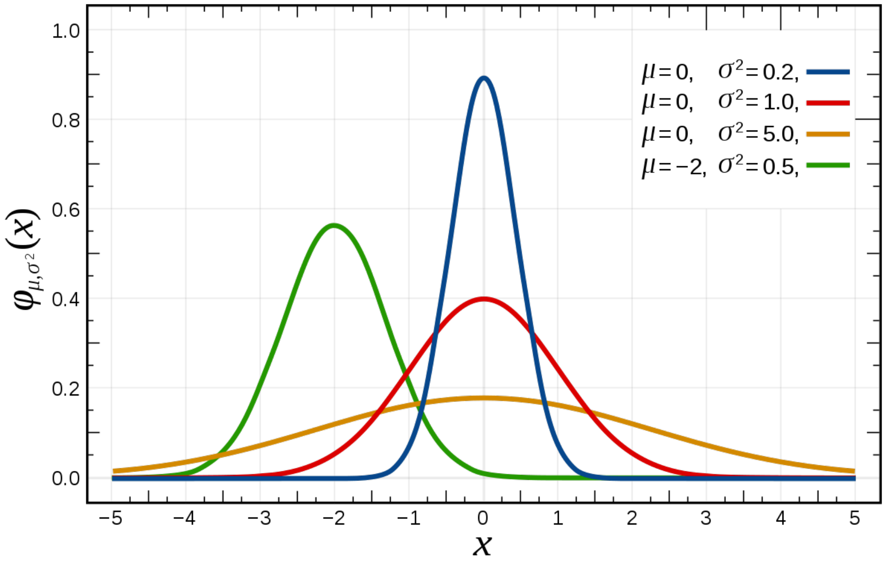
- Variance: var$(X) = \lambda$

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{k=0}^{\infty} k^2 P(X = k) = \sum_{k=0}^{\infty} k^2 \frac{1}{k!} \lambda^k e^{-\lambda} \\
&= \lambda e^{-\lambda} \left( \sum_{k=1}^{\infty} (k-1) \frac{1}{(k-1)!} \lambda^{k-1} + \sum_{k=1}^{\infty} \frac{1}{(k-1)!} \lambda^{k-1} \right) \\
&= \lambda e^{-\lambda} \left( \lambda \sum_{k=2}^{\infty} \frac{1}{(k-2)!} \lambda^{k-2} + \sum_{k=1}^{\infty} \frac{1}{(k-1)!} \lambda^{k-1} \right) \\
&= \lambda e^{-\lambda} (\lambda e^{\lambda} + e^{\lambda}) = \lambda^2 + \lambda
\end{aligned}$$

$$\text{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

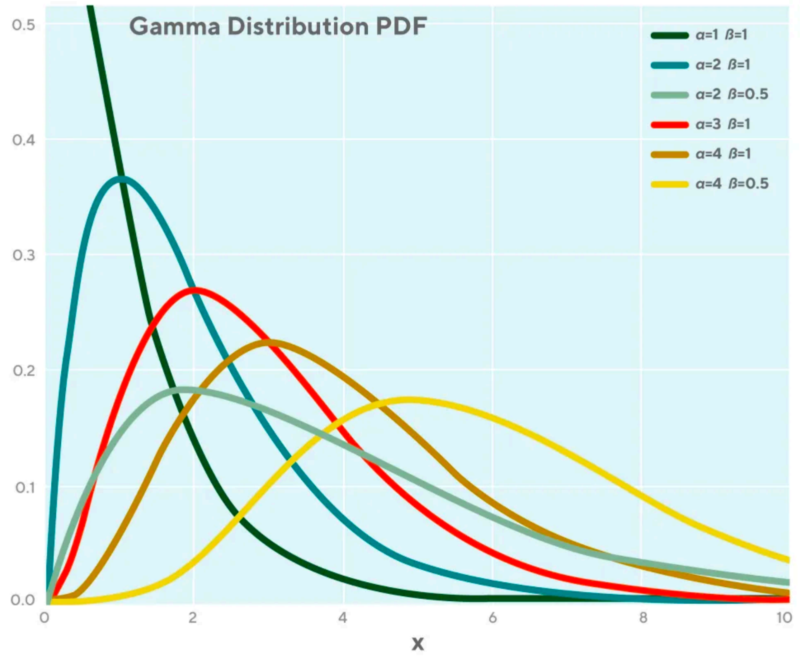
# Common distributions: Normal

- $X \sim \mathcal{N}(\mu, \sigma^2)$
- $p(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\dfrac{(x-\mu)^2}{2\sigma^2}\}, x \in \mathbb{R}$
- MGF: $M_X(t) = \exp\{\mu t + \dfrac{\sigma^2 t^2}{2}\}$
- $\mathbb{E}(x) = \mu$, $\mathrm{Var}(x) = \sigma^2$
- If $Z \sim \mathcal{N}(0,1)$, and $X = \mu + \sigma Z$, then $X \sim \mathcal{N}(\mu, \sigma^2)$.



# Common distributions: Gamma

- $X \sim \Gamma(\alpha, \beta)$
- $p_X(x) = \dfrac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, x > 0,\ x \in \mathbb{R}$
- $\Gamma(\alpha) = \displaystyle\int_0^\infty \frac{1}{\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx$
- Important statistical distribution $\chi_p^2 = \Gamma(\frac{p}{2}, 2)$
- $\chi_p^2 = \displaystyle\sum_{i=1}^{p} X_i^2$, where $X_i \sim \mathcal{N}(0,1)$, i.i.d.



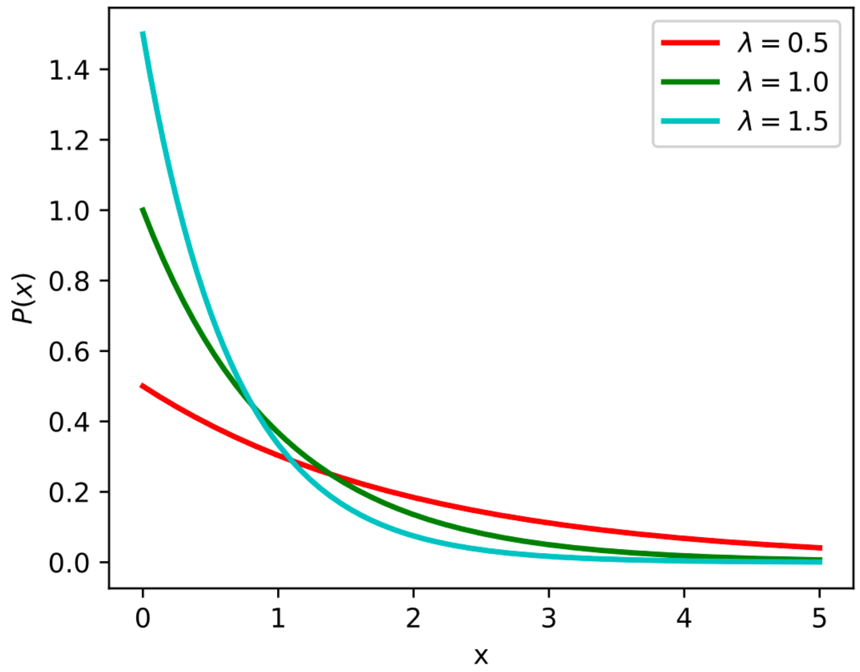
# Common distributions: Exponential

- $X \sim \exp(\lambda)$
- $p_X(x) = \lambda e^{-\lambda x}, x \in \mathbb{R}, x > 0$
- $\exp(\lambda) = \Gamma(1, 1/\lambda)$
- Mean and variance of the exponential:
- $$\mathbb{E}[T] = \int t \cdot f_T(t)dt = \int_0^\infty t\lambda e^{-\lambda t}dt$$

$$= -te^{-\lambda t}|_0^\infty + \int_0^\infty e^{-\lambda t}dt = \frac{1}{\lambda}$$

$$\mathbb{E}[T^2] = \int t^2 \cdot f_T(t)dt = \int_0^\infty t^2\lambda e^{-\lambda t}dt$$

$$= -t^2e^{-\lambda t}|_0^\infty + \int_0^\infty 2te^{-\lambda t}dt = \frac{2}{\lambda^2}$$




var$(T) = \mathbb{E}[T^2] - (\mathbb{E}[T])^2 = \frac{1}{\lambda^2}$

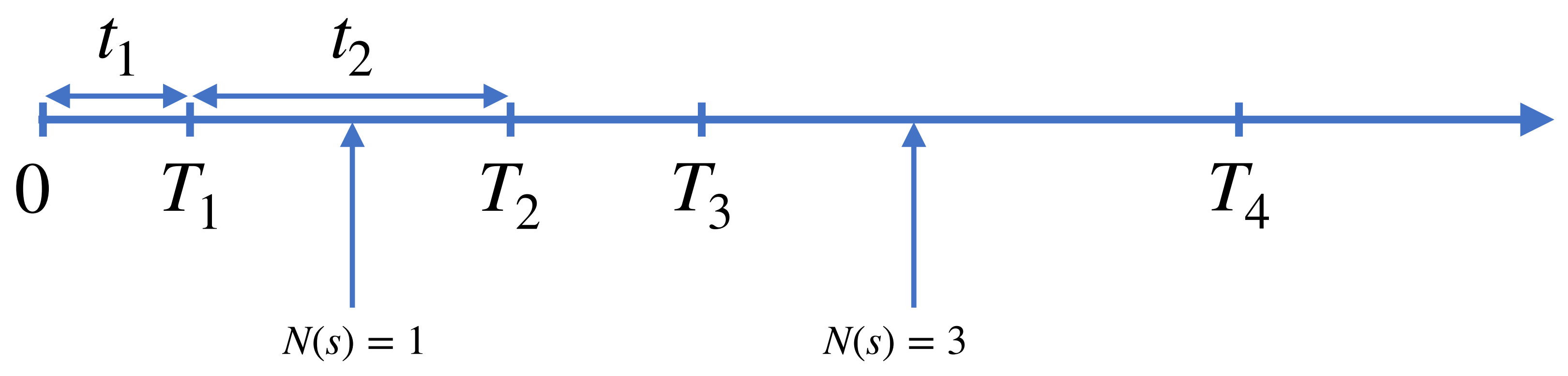
# Common distributions: Exponential

- Memoryless property of exponential
- In words:
  - Suppose the waiting time for a spike to arrive is exponentially distributed. If I've been waiting for *t* seconds, then the probability that I must wait *s* more seconds is the same as if I hadn't waited at all.
- With Math:

$$P(T > t + s \mid T > t) = P(T > s)$$

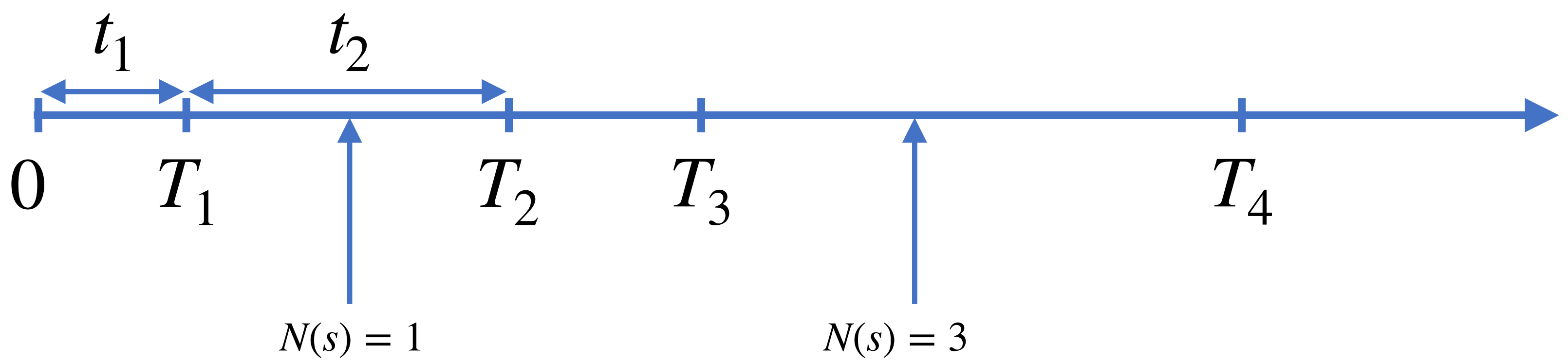
# Modeling spike trains using Poisson process

- Let $t_1, t_2, \ldots$ be independent exponential random variables with parameter $\lambda$. Let $T_n = t_1 + t_2 + \ldots + t_n$, for $n \geq 1$, $T_0 = 0$.
- Define $N(s) = \max\{n : T_n \leq s\}$
- $N(s)$ is a Poisson process



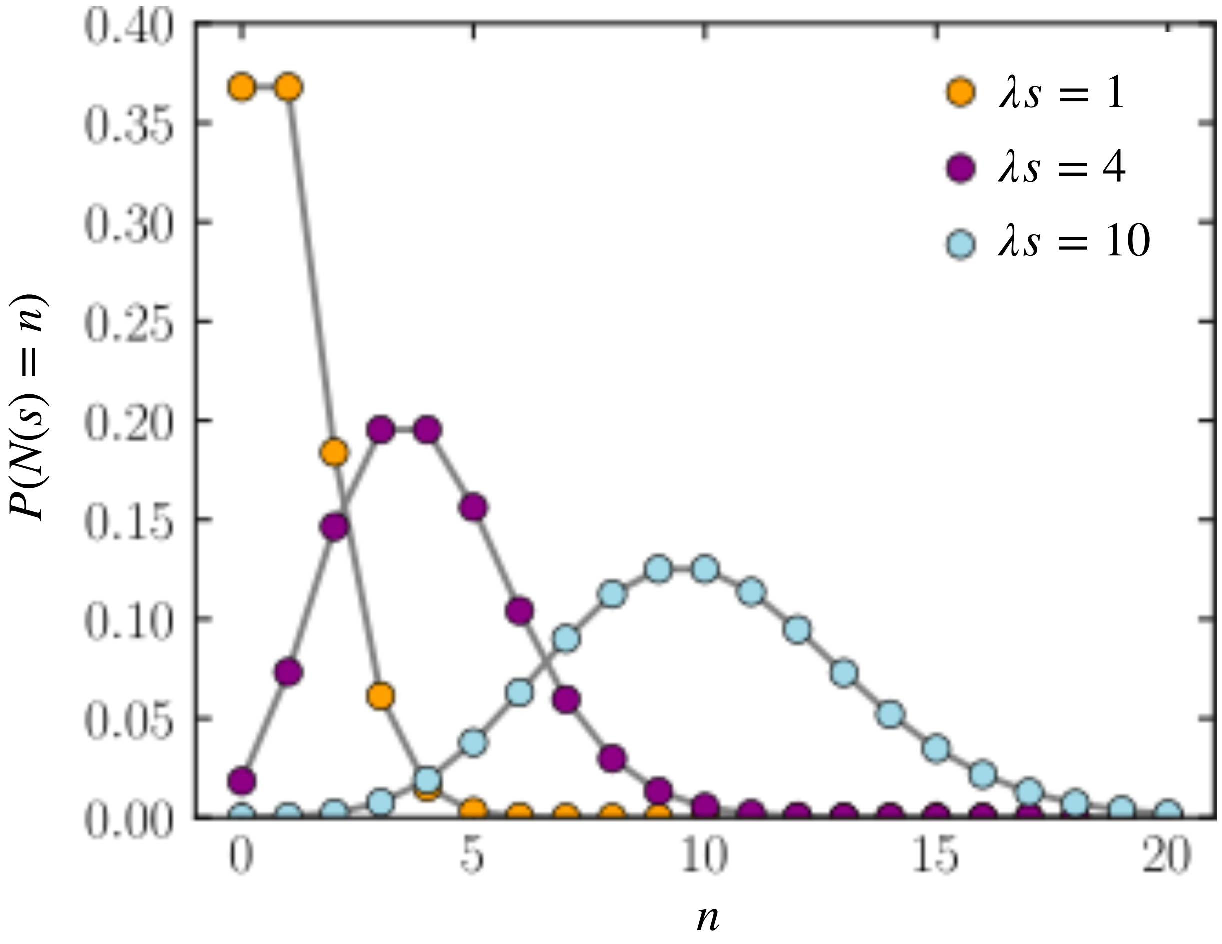
# Modeling spike trains using Poisson process

- If a Poisson process is used to model a spike train, then:
- $t_n$ is the n-th interstice interval (ISI)
- $T_n$ is the time at which the n-th spike occurs
- $N(s)$ is the number of spikes by time $s$, $\lambda$ is the neuron's firing rate
- $N(s)$ has a Poisson distribution with mean $\lambda s$.



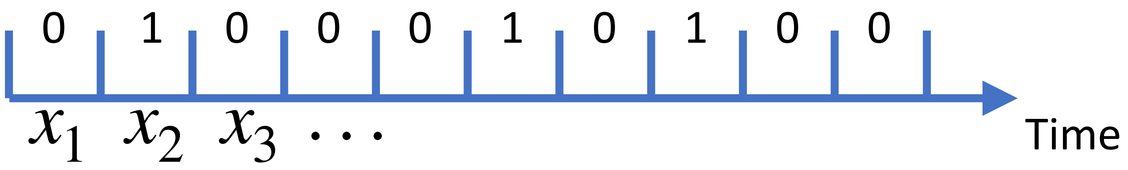
# Properties of the Poisson process

- What does a Poisson distribution look like?

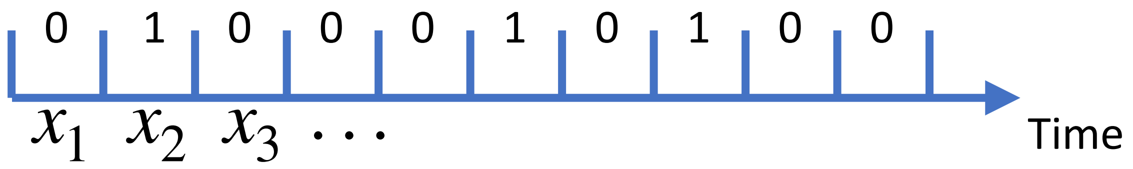


# Another view of the Poisson process

- So far, we have derived the Poisson process using i.i.d. exponential ISI's.
- Another very useful way of thinking about the Poisson process is using the <u>Bernoulli process</u>. The Poisson process is the continuous-time limit of the Bernoulli process, which is defined in discrete time.



# Bernoulli process




- $n$ is the number of discrete time steps
- $p$ is the probability of spiking at each time step
- At each time step, flip a coin to decide whether the neuron spikes (1) or not (0). The coin flips are independent of each other.
- At the i-th time step, $X_i \sim$ Bernoulli($p$), i.i.d.

$$X_i = \begin{cases} 1 & \text{w.p.} \quad p \\ 0 & \text{w.p.} \quad 1-p \end{cases}$$

# Bernoulli process

- Let $S_n$ be the number of spikes up to and including the n-th timestep

$$S_n = \sum_{i=1}^{n} X_i,\ S_n \sim \text{Bernoulli}(n, p),$$

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$\mathbb{E}[S_n] = np$ , $\Rightarrow$ we expect to see *np* spikes in *n* time steps

- As $n \to \infty$ and $p \to 0$, the Bernoulli process becomes the Poisson process, where

$$np = \lambda s$$

Mean spike count for Bernoulli process in *n* time steps

Mean spike count for Poisson process in window of duration *s*

# Bernoulli process

- What is the probability that Poisson process gives a spike in a small time window of duration $\delta$
- The number of spikes in this window is $\sim$ Poisson$(\lambda\delta)$

$$P(0 \text{ spikes in } [t, t+\delta]) = e^{-\lambda\delta} = 1 - \lambda\delta + O(\delta^2)$$

$$P(1 \text{ spikes in } [t, t+\delta]) = e^{-\lambda\delta}\lambda\delta = \lambda\delta + O(\delta^2)$$

$$P(1 \text{ spikes in } [t, t+\delta]) = O(\delta^2)$$

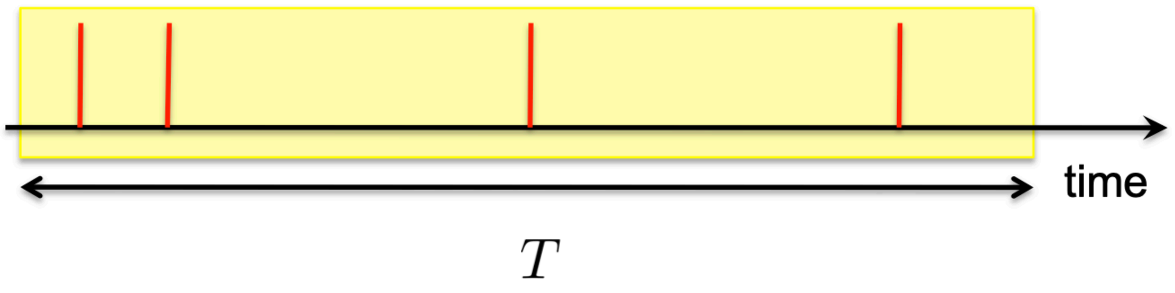
When $\delta$ is small, $O(\delta^2) \to 0$

Whether or not the neuron spikes in this window can be determined with a coin flip, where the probability of a spike is $\lambda\delta$.

# Comparison with Data

- Poisson process is simple and useful, but does it match data variability?
- Let *X* be the spike count in a bin of duration *T*

$$X \sim \text{Poisson}(\lambda T)$$



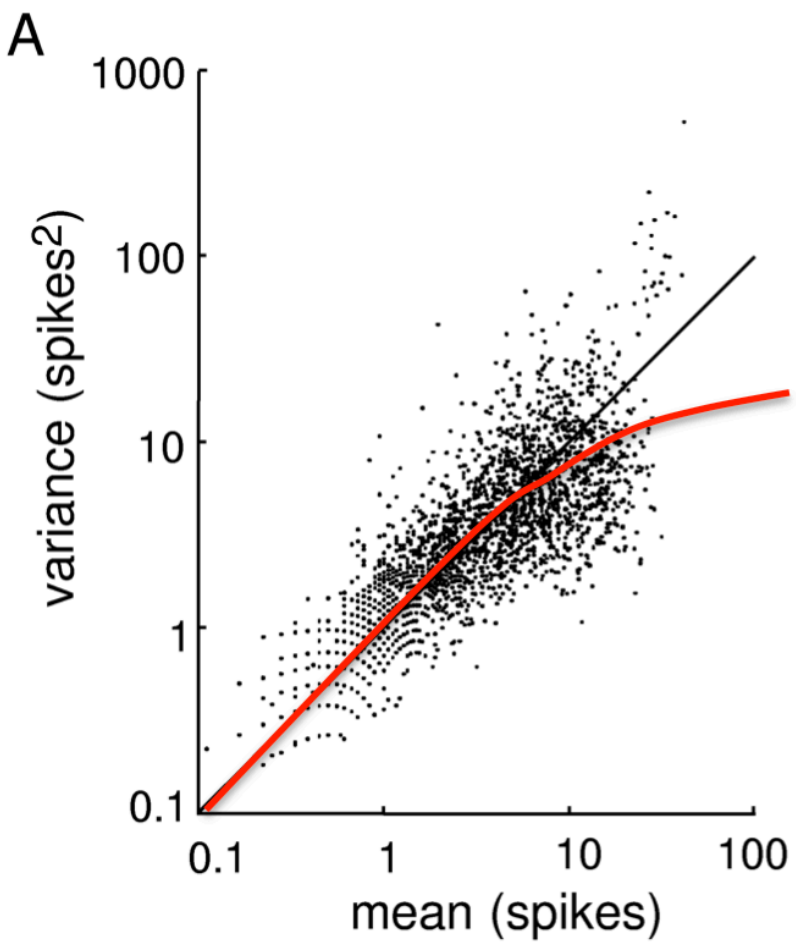

$$E[X] = \lambda T$$

$$\text{var}(X) = \lambda T$$

$$\text{Fano factor} = \frac{\text{var}(X)}{E[X]} = 1$$

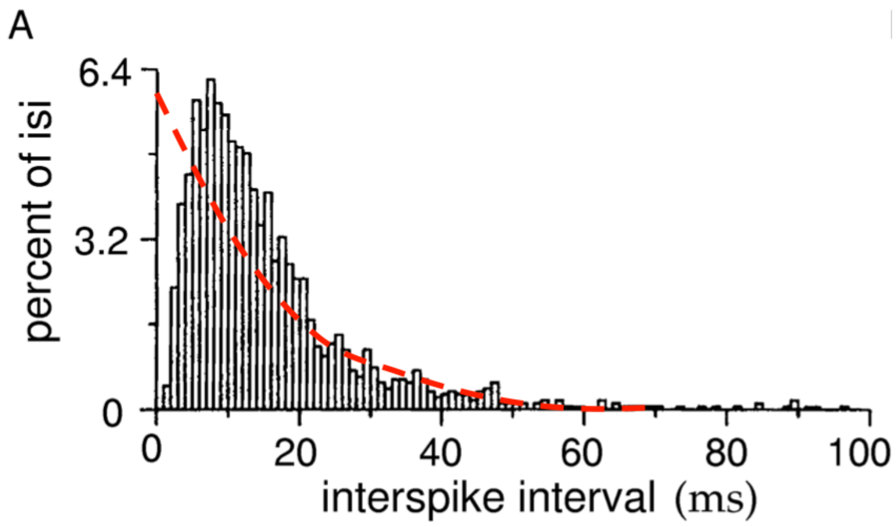
# Example from Primate Medial Temporal (MT) Area

- This is an example where a Poisson distribution models the counts well.
- Typically, fits aren't this good with real data.
- Refractory period can lead to more regular spiking (i.e. lower variance) at higher firing rates than would be predicted by Poisson.



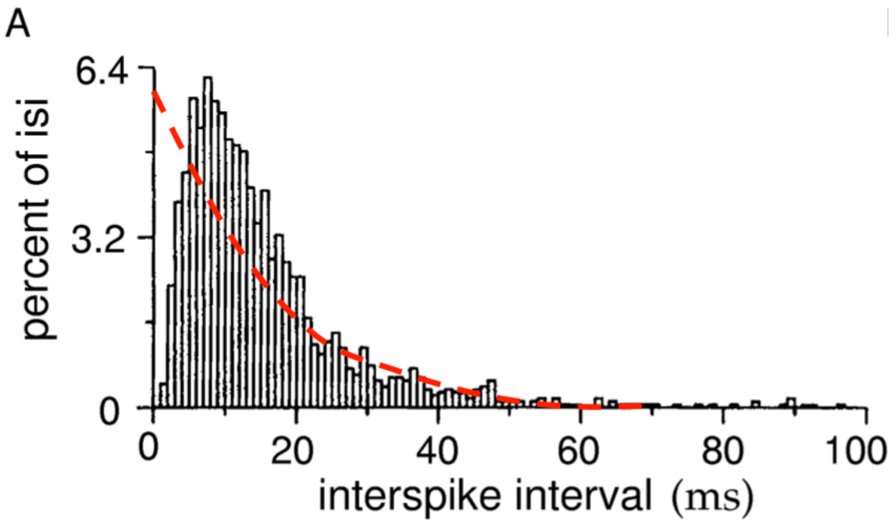
# Inter-spike Interval (ISI) Distribution

- Poisson process predicts exponentially-distributed ISI’s (red curve).
- Example from primate MT area:




- This ISI distribution is very typical of real neural data.
- Why are ISI’s not exponentially distributed for real neural data?

# Why are ISI's not exponentially distributed?

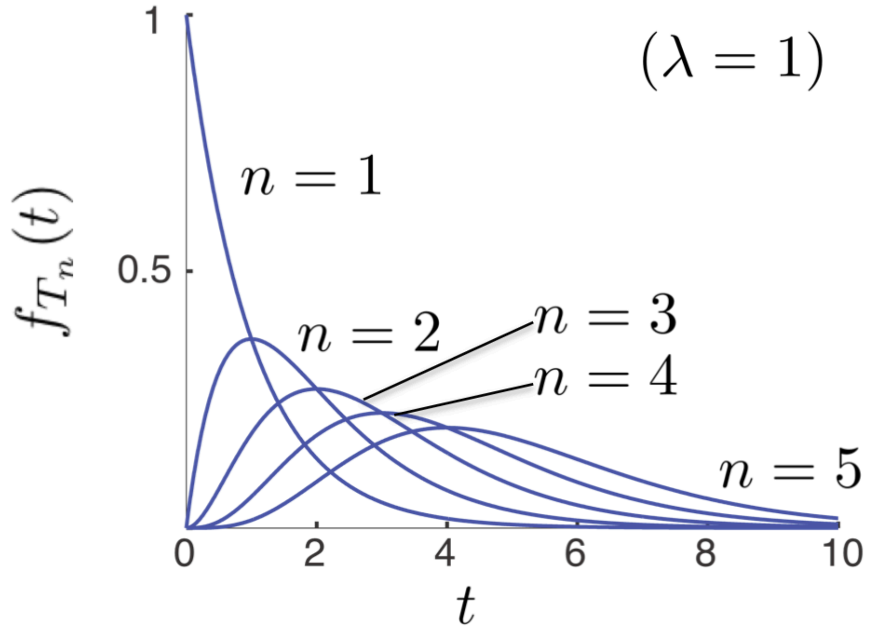



1) Real data have few or no short ISI's due to refractory period.

2) Firing rates typically vary over time.

- What is a better model for ISI's?

# Gamma Distribution




$n$ is the order of the Gamma distribution

$n = 1$ is the exponential distribution

ISI's are better modeled using Gamma distribution with $n > 1$

# Gamma Distribution

Let $T_n = t_1 + \ldots + t_n$, where $t_1, \ldots, t_n \sim \exp(\lambda)$ i.i.d.

$T_n$ is an $n$th order Gamma random variable

$$f_{T_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \qquad \text{for } \lambda > 0 \text{ and } t > 0.$$

Intuition: The larger the order $n$ (i.e., the more exponentially-distributed random variables are added together), the fewer small ISI's will appear in the Gamma distribution.

# Coefficient of Variation ($C_V$) of ISI's

Let $t \sim \exp(\lambda)$,

$$E[t] = \frac{1}{\lambda} \qquad \mathrm{var}(t) = \frac{1}{\lambda^2}$$

$$C_v = \frac{\sqrt{\mathrm{var}(t)}}{E[t]} = 1$$

# Coefficient of Variation ($C_V$) of ISI's