Lecture 25: if $N_1 \sim \text{Poisson} (\mu)$, $N_2 \sim \text{Poisson} (\lambda)$ and $N = N_1 + N_2$

we already know that $E(N) = E(N_1) + E(N_2) = \mu + \lambda$. If $N_1$ & $N_2$ are independent we also have $\text{Var}(N) = \text{Var}(N_1) + \text{Var}(N_2) = \mu + \lambda$.

*Theorem:* In fact, Poisson r.v.'s have the following remarkable property:

if $N_1 \sim \text{Poisson} (\mu)$, $N_2 \sim \text{Poisson} (\lambda)$ are independent, and

$N = N_1 + N_2$

then: $N \sim \text{Poisson} (\mu + \lambda)$, i.e. $P(N = k) = e^{-(\mu+\lambda)}(\mu + \lambda)^k / k!$, $k=0,1,2,...$

*Proof:* $P(N = k) = P(N_1 + N_2 = k) = \text{average of conditional probabilities}$

$= \sum_{j=0}^{k} P(N_1 + N_2 = k \mid N_1 = j) P(N_1 = j)$

$= \sum_{j=0}^{k} P(N_2 = k-j \mid N_1 = j) P(N_1 = j)$

$= \sum_{j=0}^{k} P(N_2 = k-j) P(N_1 = j)$

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

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

$= \sum_{j=0}^{k} \frac{k!}{j!(n-j)!} \left( \frac{\mu}{\mu + \lambda} \right)^j \left( \frac{\lambda}{\mu + \lambda} \right)^{k-j}$

$= \left( \frac{\mu}{\mu + \lambda} + \frac{\lambda}{\mu + \lambda} \right)^k = 1$

$= e^{-(\mu + \lambda)} \frac{(\mu + \lambda)^k}{k!}$
More in general, if \( N_1, N_2, \ldots, N \) are independent Poisson r.v.'s with parameters \( \mu_1, \mu_2, \ldots, \mu \), then \( N = N_1 + N_2 + \ldots + N \) is Poisson \( (\mu) \) with \( \mu = \mu_1 + \mu_2 + \ldots + \mu \).

Example: if I play \( m \) games, each with small probability of success \( p_1, p_2, p_3, \ldots, p_m \) (e.g. 1st game: success is getting "0" at the roulette, \( p_1 = \frac{1}{38} \), 2nd game: success is rolling two 6's simultaneously, \( p_2 = \frac{1}{36} \), etc.)

and I play them \( n_1, n_2, \ldots, n_m \) times respectively, and I call: \( N_1 = \# \) of successes in first game \( \sim \) Poisson \( (n_1 p_1) \)
\( N_2 = \# \) of successes in 2nd game \( \sim \) Poisson \( (n_2 p_2) \)

then: \( N = \sum_{i=1}^{m} N_i = \# \) of successes \( \sim \) Poisson \( (n_1 p_1 + \ldots + n_m p_m) \).

Remark: if \( N \sim \text{Poisson} (\mu) \), with \( \mu \) fairly large, we may think of \( N \) as the sum of \( n \) independent Poisson r.v.'s \( N_1, \ldots, N_n \)
\( N = N_1 + \ldots + N_n \)
each with parameter \( \mu_i = \frac{\mu}{n} : E(N_i) = \frac{\mu}{n}, \text{SD}(N_i) = \sqrt{\frac{\mu}{n}} \).
So by the Central Limit Theorem, we have that \( N \) may be approximated by a Gaussian with mean \( n \frac{\mu}{n} = \mu \) & \( \sigma = \sqrt{n} \sqrt{\frac{\mu}{n}} = \sqrt{\mu} \).
(Again, see histograms).

---

For us, this is the end of Chapter 3.
For Midterm #2, up to here.
Probability densities: they are "continuous distributions" (defined on a discrete set.)

A "discrete random variable" has a "discrete distribution" $P(Y = y)$, represented by a histogram where the height of the bar above $x$ is proportional to $P(Y = y)$ (in fact, equal to $P(Y = y)$ when the width of the basis of each bar is equal to one):

$$P(a \leq Y \leq b) = \sum_{a \leq y \leq b} P(Y = y)$$

A "continuous random variable" $X$ has a "probability density" (or "continuous distribution") $f(x)$ defined for $x \in \mathbb{R}$ (a continuous set), which is a function such that:

$$P(a \leq X \leq b) = \int_{a}^{b} f(x) \, dx$$

These are used to describe quantities that are continuous in nature, such as height, temperature, electric potential, chemical concentration (as opposed to things that are discrete in nature, such as # of successes in repeated and independent trials, # of red cards out of 5 in a deck of 52, age of a person, lowest # I get when rolling 4 dice, # of students in a class, etc.)
Remarks: for a continuous r.v. \( X \) with prob. density \( f(x) \):

- \( P(X = b) = \int_a^b f(x) \, dx = 0 \), for all \( b \in \mathbb{R} \)
  (the probability of, say, choosing a student at random whose height is exactly \( b = 6'2'' \) is 0).

- \( f(x) \) is not a probability, but a probability density:
  take "\( \Delta x \)" sufficiently small, we have
  \[
  P(X \in [a, a+\Delta x]) = \frac{\int_a^{a+\Delta x} f(x) \, dx}{\Delta x} \approx f(a) \Delta x.
  \]

  So \( f(a) = \frac{P(X \in [a, a+\Delta x])}{\Delta x} \),
  and that's why we say it's a "density": it's measured in probability per unit of measurement of \( \Delta x \).

Example: if \( X \) describes height (measured in meters) of a population, the unit of measurement for its density \( f(x) \) is \( m^{-1} \).

More remarks:
- we have \( f(x) \geq 0 \) for all \( x \), otherwise we could have negative probabilities.

  Since we want \( P(-\infty < X < \infty) = 1 \),
  we must have \( \int_{-\infty}^{\infty} f(x) \, dx = 1 \)

  \( f(x) \) is a probability density function (PDF).

- However, note that (while for a discrete probability distribution \( P(X = x) \leq 1 \)), for a continuous density we can have \( f(x) > 0 \), as long as (*) holds.
Most concepts that apply to discrete r.v.'s also apply to continuous r.v.'s:

**DISCRETE r.v. Y:**

\[
E(Y) = \sum_y y P(Y = y)
\]

**mean:**

\[
E(Y^2) = \sum_y y^2 P(Y = y)
\]

**2nd moment:**

**CONTINUOUS r.v. X:**

\[
\mu = E(X) = \int_{-\infty}^{\infty} x f(x) \, dx
\]

\[
E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) \, dx
\]

**Similarly:**

\[
E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) \, dx
\]

**Variance:**

\[
\text{Var}(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2
\]

**can prove**

\[
\text{SD}(X) = \sqrt{\text{Var}(X)}
\]

**Independence:** two continuous random variables X and Y are called independent if for any pair of intervals \(A \subseteq \mathbb{R} \) and \(B \subseteq \mathbb{R} \) we have that

\[
P(X \in A, Y \in B) = P(X \in A)P(Y \in B).
\]

(we shall see later on that this translates into the factorization of the "joint probability density" into the marginals:

\[
f_{X,Y}(x,y) = f_X(x) f_Y(y).
\]

**Finally, for continuous random variables we have that:**

* Markov's inequality holds
* Chebychev's inequality holds
* The Law of Averages (a.k.a. the Law of Large Numbers) holds
* The Central Limit Theorem holds too.