Last time: examples of discrete-valued random variables

- Binomial
- Uniform
- Poisson
- Geometric

Geometric distribution for a r.v. $X$

$$\text{Prob}(X = j) = (1-q)q^j \quad \text{for } j \geq 0$$

and $j$ an integer ($j \in \mathbb{Z}_{\geq 0}$)

**Note:** Usually models the probability that the first "success" in a sequence of independent, identically distributed "bernoulli trials" happens at the trial $j$.

$q = \text{probability of failure}

If we start with trial # 0.

If we want to start with trial # 1

$$\text{Prob}(X = j) = (1-q)q^{j-1} \quad \text{for } j \in \mathbb{N}$$
Application in genome sequencing

\[ \begin{align*}
    C & \quad T & \quad T & \quad A & \quad G & \quad C & \quad dT \\
    & \quad A & \quad T & \quad C & \quad A & \quad C & \quad T & \quad A
\end{align*} \]

\[ \begin{align*}
    T & \quad G & \quad C & \quad C & \quad A & \quad tT
\end{align*} \]

Several discrete-valued random variables put this in same framework by letting the state space be vector space.

Example: Rainfall = \( X_1 \in \mathbb{R}_{\geq 0} \)

Mosquito Population: \( X_2 \in \mathbb{R}_{\geq 0} \)

\# cases West Nile disease = \( X_3 \in \mathbb{R}_{\geq 0} \)

\[
    \vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}
\]

State space \( \vec{X} \in S = \mathbb{R}_{\geq 0}^3 \times \mathbb{R}_{\geq 0}^2 \)
But often care about how the random variables are related to each other.

Two r.v.'s $X$ and $Y$,

- said to be independent if there is no statistical connection between them

$$\text{Prob}(X \in A \text{ and } Y \in B) = \text{Prob}(X \in A) \times \text{Prob}(Y \in B)$$

for any sets $A \subseteq S_X$ : state space for $X$

$B \subseteq S_Y$ : state space for $Y$

More simply for discrete cases:

$$\text{Prob}(X = x \text{ and } Y = y) = \text{Prob}(X = x) \times \text{Prob}(Y = y)$$
How quantity dependence of one random variable on another?

Covariance:
\[
\text{Cov}(X, Y) = \langle (X - \mu_X)(Y - \mu_Y) \rangle
\]
where \( \mu_X = \langle X \rangle \), \( \mu_Y = \langle Y \rangle \)

\( = 0 \) for independent \( X, Y \)

\[
\text{Fact: } \langle f(X)g(Y) \rangle = \langle f(X) \rangle \langle g(Y) \rangle
\]
if \( X \) and \( Y \) are independent

Proof: \( \langle f(X)g(Y) \rangle = \sum_{x \in S_x} \sum_{y \in S_y} f(x)g(y) \Pr(X=x, Y=y) \)

\( > 0 \) if \( X \) and \( Y \) are "positively correlated"
\[ X \uparrow \Rightarrow Y \uparrow \]
\[ X \downarrow \Rightarrow Y \downarrow \]

\( < 0 \) if \( X \) and \( Y \) are "negatively correlated"
\[ X \uparrow \Rightarrow Y \downarrow \]
\[ X \downarrow \Rightarrow Y \uparrow \]
Conditional expectation:

\[
\langle X \mid Y = y \rangle = E(X \mid Y = y) = \sum_{x \in S_X} x \cdot \text{Prob}(X = x \mid Y = y)
\]

Some other terminology:

If we have a collection \( \{ X_j \}_{j=1}^N \) of random variables,

\[
P_N(x_1, x_2, \ldots, x_N) = \text{Prob}(X_1 = x_1 \text{ and } X_2 = x_2 \ldots \text{ and } X_N = x_N)
\]

Joint probability distribution: keep track of events involving 2 or more rvs,

Marginal probability distribution:
Condenses the information down to what's necessary to study 1 r.v. at a time,

\[
p_{N,j}(x) = \text{Prob}(\bar{X}_j = x) = \sum_{x_j \in S_{X_j}} p_N(x_1, x_2, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_N)
\]
How strong is the correlation between 2 rvs?

Correlation coefficient

\[ \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \]

standard deviations

\[-1 \leq \rho(X, Y) \leq 1 \]

strong correlation

More general way to relate random variables is by conditional statistics.

Conditional probability

\[ \text{Prob}(X = x \mid Y = y) = \frac{\text{Prob}(X = x \text{ and } Y = y)}{\text{Prob}(Y = y)} \]
Probability generating function and characteristic function - useful tools for doing complicated calculations in prob. theory

Consider a random variable $X$ with state space $S = \mathbb{Z}_{\geq 0}$

Probability generating function

$$G_X(s) = \mathbb{E}[s^X] = \sum_{j=0}^{\infty} s^j \text{Prob}(X=j)$$

Characteristic function

$$\Phi_X(k) = \mathbb{E}[e^{ikX}] = \sum_{j \in S} e^{ikj} \text{Prob}(X=j)$$

Knowing either of these is equivalent to knowing probability distribution of $X$.

(Knowing $p_j = \text{Prob}(X=j)$ for $j \in S$)

$$G_X(s) = \Phi_X(-i \ln s), \quad \Phi_X(k) = G_X(e^{ik})$$
\[
\mathbf{G}_X(s) = \sum_{j=0}^{\infty} \mathbf{G}_X^{(j)} s^j \text{ Prob}(X = j)
\]

\[
\text{Prob}(X = j) = \left. \frac{1}{j!} \left( \frac{d^j}{ds^j} \right) \mathbf{G}_X(s) \right|_{s=0}
\]

Another interesting relationship:

\[
\langle X^n \rangle = \left. \left( -i \frac{d}{dk} \right)^n \mathbf{G}_X(k) \right|_{k=0}
\]

(Direct observation from \( \mathbf{G}_X(k) = \langle e^{ikX} \rangle \))

(So \( \mathbf{G}_X(k) \) is sometimes called "moment generating function")

Cumulant generating function:

\[
\mathbf{M}_{X,N} = \langle \langle X^N \rangle \rangle = \left( -i \frac{d}{dk} \right)^n \mathbf{G}_X(k)
\]

where \( N \) th order cumulant of r.v. \( X \)

\[
\tilde{\mathbf{G}}_X(k) = \ln \mathbf{G}_X(k) \] is the cumulant generating function
Cumulants are a more efficient way of organizing information contained in moments.

\[
\begin{align*}
\langle X \rangle \\
\langle X^2 \rangle \\
\sigma_X^2 &= \langle X^2 \rangle - \langle X \rangle^2
\end{align*}
\]

\[
\begin{align*}
M_{X,1} &= \langle X \rangle \\
M_{X,2} &= \langle X^2 \rangle - \langle X \rangle^2 = \sigma_X^2 \\
M_{X,3} &= \langle X^3 \rangle - 3\langle X \rangle \langle X^2 \rangle - \langle X \rangle^3
\end{align*}
\]

One way this manifests itself is joint moments.

\[
\begin{align*}
\langle X^2 Y^2 \rangle
\end{align*}
\]

Suppose \( \langle X \rangle = 0, \langle Y \rangle > 0 \).

Then one may want to look at a moment like this to look at relations in variables between \( X \) and \( Y \).

But even if \( X \) and \( Y \) don't have anything to do with each other,

\[
\begin{align*}
\langle X^2 Y^2 \rangle &= \langle X^2 \rangle \langle Y^2 \rangle \neq 0
\end{align*}
\]
But if we use cumulants...

Cumulant (joint) generating function

$$\hat{\phi}_{X,Y}(k,k') = \ln \frac{e^{ikX + ik'Y}}{e^{kX}e^{k'Y}}$$

$$M_{X,X,2,2} = \left. \left( -i \frac{\partial^2}{\partial k} \right) \left( -i \frac{\partial^2}{\partial k'} \right) \hat{\phi}_{X,Y}(k,k') \right|_{k=k'=0}$$

$$= \langle X^2 Y^2 \rangle - \langle X^2 \rangle \langle Y^2 \rangle$$

if $$\langle X \rangle = \langle Y \rangle = 0$$.

Physicist way of calculating cumulants - diagrams

"subtract off disconnected diagrams"
Back to normal generating functions.

How are these useful?
- Sometimes they're useful for calculating moments.
- Calculations involving recursive stochastic processes, and more generally stochastic processes where independent information comes in discrete chunks and in a "homogeneous" way.

Let $X_1$ and $X_2$ be independent r.v.s,

$$Y = X_1 + X_2$$

$$\Pr(Y = j) = \sum_{j' \in S} \Pr(X_1 = j' \text{ and } X_2 = j - j')$$

(independence) = \sum_{j' \in S} \Pr(X_1 = j') \Pr(X_2 = j - j')

"convolution"

Easier to work with characteristic fn or prob gen. fn.
\[ G_Y(s) = \langle S X \rangle = \langle S X_1 + X_2 \rangle = \langle S X_1, S X_2 \rangle = \langle S X_1 \rangle \langle S X_2 \rangle = G_{X_1}(s) G_{X_2}(s) \]

\[ S_N = X_1 + X_2 + \ldots + X_N \quad \text{with } X_j \text{ are i.i.d. independently distributed} \]

\[ G_{S_N}(s) = \prod_{j=1}^{N} G_{X_j}(s) = G_{X}(s)^N \]

\[ \phi_{S_N}(k) = \left( \phi_{X}(k) \right)^N \]
Example of using generating functions for calculating moments.

**Binomial distribution for** \( n \in \mathbb{Y} \)

\[
\text{Prob}(Y = j) = \binom{N}{j} p^j (1-p)^{N-j}
\]

\( 0 < p < 1 \)

\[
\mathbb{E}[Y^n] = \sum_{j=0}^{N} j^n \binom{N}{j} p^j (1-p)^{N-j}
\]

... and even for \( n = 1, 2, \ldots \)

**Probability gen. fn.**

\[
G_Y(s) = \sum_{j=0}^{N} \frac{(N)_j}{j!} p^j (1-p)^{N-j} s^j
\]

\[
\text{Prob}(Y = j) = \sum_{j=0}^{N} \binom{N}{j} (ps)^j (1-p)^{N-j}
\]

\[
G_Y(s) = (1-p + ps)^N
\]

(binomial expansion formula)
No free this is consistent with

\[ Y = X_1 + X_2 + \ldots + X_N \]

where \( X_j \) are i.i.d. \( \text{Prob}(X_j = 1) = p \)

\[ \text{Prob}(X_j = 0) = 1 - p \]

\[ G_Y(s) = \left( \frac{G_X(s)}{s} \right)^N \]

\[ G_X(s) = (1 - p)s^0 + ps \]

\[ < Y^n > = \left( -i \frac{d}{dk} \right)^n \phi_Y(k) \bigg|_{k=0} = \left( \frac{d}{d(\ln s)} \right)^N G_Y(s) \bigg|_{s=1} \]

\[ \phi_Y(k) = G_Y(e^{ik}) \]

\[ \phi_Y(k) = (1 - p + p e^{ik})^N \]

\[ < Y^n > = \left( -i \frac{d}{dk} \right)^n \phi_Y(k) \bigg|_{k=0} \]

\[ < Y > = Np \]

\[ < Y^2 > = pN + N(N-1)p^2 \]

\[ \sigma^2_Y = < Y^2 > - < Y >^2 = N(p - p^2) \]

\[ \sigma_Y = \sqrt{N(p - p^2)} \]