01/22/04

Last time: examples of discrete-valued random variables

Binomial
Uniform
Poisson
Geometric

Geometric distribution for a r.v. X

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

and j an integer

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

Usually models the probability that the first "success" in a sequence of independent, identically distributed "Bernoulli trials" happens at ~~the~~ trial j

q = probability of failure

If we start with trial # 0.

If we want to start with trial # 1

$$\text{Prob}(X = j) = (1-q)\, q^{j-1} \qquad \text{for } j \in \mathbb{N}$$

Application to genome sequencing

G T T A G C dT A

A A T C G A C T A

T G C C A

dT T

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: $X_2 \in \mathbb{Z}_{\geq 0}$

cases West Nile disease $= X_3 \in \mathbb{Z}_{\geq 0}$

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

State space

$$\vec{X} \in S = \mathbb{R}_{\geq 0} \times \mathbb{Z}^2_{\geq 0}$$

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

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

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

More simply for discrete case:

$$\text{Prob}(X = x \text{ and } Y = y) = \text{Prob}(X=x)\,\text{Prob}(Y=y)$$

How quantity dependence of one random variable on another?

Covariance:

$$\mathrm{Cov}(X, Y) \equiv \langle (X - \mu_X)(Y - \mu_Y) \rangle$$

where $\mu_X = \langle X \rangle$, $\mu_Y = \langle Y \rangle$

$= 0$ for independent X, Y

(Fact: $\langle f(X) g(Y) \rangle = \langle f(X) \rangle \langle g(Y) \rangle$ if X and Y are independent)

Proof: $\langle f(X) g(Y) \rangle = \sum_{\substack{x \in S_X \\ y \in S_Y}} f(x) g(y) \, \mathrm{Prob}\begin{pmatrix} X = x \\ \text{and} \\ Y = y \end{pmatrix}$

...

> 0 if X and Y are "positively correlated"

$X \uparrow \Rightarrow Y \uparrow \qquad X \downarrow \Rightarrow Y \downarrow$

< 0 if X and Y are "negatively correlated"

$X \uparrow \Rightarrow Y \downarrow \qquad X \downarrow \Rightarrow Y \uparrow$

Conditional expectation:

$$\langle X \mid Y = y \rangle \equiv \mathbb{E}(X \mid Y = y)$$

$$= \sum_{x \in S_X} x \, \text{Prob}(X = x \mid Y = y)$$

Some other terminology:

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

Joint probability distribution: keep track of events involving 2 or more r.v.'s.

$$p_N(x_1, x_2, \ldots, x_N) = \text{Prob}(X_1 = x_1 \text{ and } X_2 = x_2 \ldots \text{ and } X_N = x_N)$$

Marginal probability distribution:

- condenses the information down to what's necessary to study 1 r.v. at a time.

$$p_N^{(j)}(x) \equiv \text{Prob}(X_j = x) = \sum_{\substack{x_1 \in S \\ x_2 \in S \\ \vdots \\ x_{j-1} \in S \\ x_{j+1} \in S \\ \vdots \\ x_N \in S}} p_N(x_1, x_2, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_N)$$

How strong is the correlation between 2 rv's?

Correlation coefficient

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \ \sigma_Y}$$

standard deviations

$$-1 \leq \rho(X,Y) \leq 1$$

strong correlation

strong correlation

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

Conditional probability

$$Prob(X=x \mid Y=y) = \frac{Prob(X=x \text{ and } Y=y)}{Prob(Y=y)}$$

Probability generating function and characteristic function

Karlin Taylor Sec. 1.1

- 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

$$G_X(s) = \langle s^X \rangle = \sum_{j=0}^{\infty} s^j \, \mathrm{Prob}(X=j)$$

Characteristic function

$$\phi_X(k) = \langle e^{ikX} \rangle = \sum_{j \in S} e^{ikj} \, \mathrm{Prob}(X=j)$$

Knowing either of these is equivalent to knowing probability distribution of X. (Knowing $p_j = \mathrm{Prob}(X=j)$ for $j \in S$),

$$G_X(s) = \phi_X(-i \ln s), \quad \phi_X(k) = G_X(e^{ik})$$

$$G_X(s) = \sum_{j=0} s^j \, \text{Prob}(X = j)$$

$$\text{Prob}(X = j) = \frac{1}{j!} \left(\frac{d^j}{ds^j}\right) G_X(s) \Bigg|_{s=0}$$

Another interesting relationship:

$$\langle X^n \rangle = \left(-i \frac{d}{dk}\right)^n \phi_X(k) \Bigg|_{k=0}$$

(Direct observation from $\phi_X(k) = \langle e^{ikX} \rangle$)

(So $\phi_X(k)$ is sometimes called "moment generating function")

Cumulant generating function:

$$M_{X,N} \equiv \langle\langle X^N \rangle\rangle = \left(-i \frac{d}{dk}\right)^n \tilde{\phi}_X(k) \quad \text{where}$$

N th order cumulant of r.v. X

$\hat{\phi}_X(k) = \ln \phi_X(k)$ is the cumulant generating fn

Cumulants are a more efficient way of organizing information contained in moments.

$\langle X \rangle$ $\qquad\qquad$ $\mu_{X,1} = \langle X \rangle$

$\langle X^2 \rangle$ $\qquad\qquad$ $\mu_{X,2} = \langle X^2 \rangle - \langle X \rangle^2$

$\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$ $\qquad\qquad$ $= \sigma_X^2$

σ_X $\qquad\qquad$ $\mu_{X,3} = \langle X^3 \rangle - 3\langle X \rangle \langle X^2 \rangle - \langle X \rangle^3$

One way this manifests itself is joint moments.

$\langle X^2 Y^2 \rangle$ $\qquad$ 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 volatility between X and Y.

But even if X and Y don't have anything to do w/ each other

$$\langle X^2 Y^2 \rangle = \langle X^2 \rangle \langle Y^2 \rangle \neq 0.$$

But if we use cumulants...

Cumulant (joint) generating function

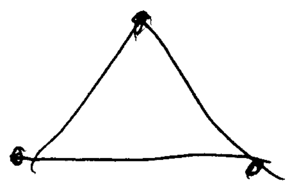
$$\hat{\phi}_{X,Y}(k, k') = \ln \langle e^{ikX + ik'Y} \rangle$$

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

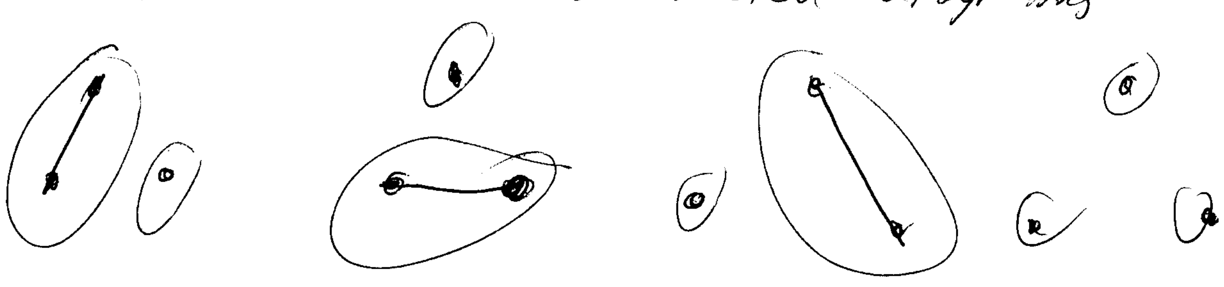
$$= \langle X^2 Y^2 \rangle - \langle X^2 \rangle \langle Y^2 \rangle$$

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
  - why?

    Let X_1 and X_2 be independent rvs,

    $$Y = X_1 + X_2$$

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

$$(\text{independence}) = \sum_{j' \in S} \mathrm{Prob}(X_1 = j')\,\mathrm{Prob}(X_2 = j - j')$$

"convolution"

Easier to work with characteristic fn or prob. gen. fn.

$$G_Y(s) = \langle s^Y \rangle = \langle s^{X_1+X_2} \rangle = \langle s^{X_1} s^{X_2} \rangle$$

$$= \langle s^{X_1} \rangle \langle s^{X_2} \rangle = G_{X_1}(s)\, G_{X_2}(s)$$

$S_N = X_1 + X_2 + \dots + X_N$ with X_j are i. i. d.

independent identically distributed

$$G_{S_N}(s) \overset{\text{independent}}{=} \prod_{j=1}^{N} G_{X_j}(s) \overset{\text{identical}}{=} \left(G_X(s)\right)^N$$

$$\phi_{S_N}(k) = \left(\phi_X(k)\right)^N$$

Homework due 02/02/04

Example of using generating functions for calculating moments.

Binomial distribution for r.v. Y

$$\mathrm{Prob}(Y=j) = \binom{N}{j} p^j (1-p)^{N-j}$$

$$0<p<1$$

$$\langle Y^n \rangle = ? \qquad \sum_{j=0}^{N} j^n \binom{N}{j} p^j (1-p)^{N-j}$$

... a pain even for $n=1,2$.

Probability gen fn:

$$G_Y(s) = \sum_{j=0}^{N} \underbrace{\binom{N}{j} p^j (1-p)^{N-j}}_{\mathrm{Prob}(Y=j)} s^j$$

$$= \sum_{j=0}^{N} \binom{N}{j} (ps)^j (1-p)^{N-j}$$

$$G_Y(s) = (1-p+ps)^N$$ (binomial expansion formula)

Notice this is consistent with

$$Y = X_1 + X_2 + \dots + X_N$$

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

$\mathrm{Prob}(X_j = 0) = 1 - p$

$$G_Y(s) = \left(G_X(s)\right)^N \qquad G_X(s) = (1-p)s^0 + ps^1 = 1 - p + ps$$

$$\langle Y^n \rangle = \left(-i\frac{d}{dk}\right)^n \phi_Y(k)\Big|_{k=0} = \left(\frac{d}{d\ln s}\right)^N G_Y(s)\Big|_{s=1}$$

$$\cancel{\phi_Y(k) = G_Y(-i\ln s)}$$

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

$$\phi_Y(k) = \left(1 - p + pe^{ik}\right)^N$$

$$\langle Y^n \rangle = \left(-i\frac{d}{dk}\right)^n \phi_Y(k)\Big|_{k=0}$$

$$\langle Y \rangle = Np$$

$$\langle Y^2 \rangle = pN + N(N-1)p^2$$

$$\sigma_Y^2 = \langle Y^2 \rangle - \langle Y \rangle^2 = N(p - p^2)$$

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