Last time (Session 16)
Poisson r.v.'s
Started Poisson scatters

This time (Session 17)
More Poisson scatters
(Cancer clusters)

Information on Exam 1
Geometric mean $\sqrt{p}$, variance $\frac{1}{p}$

For $X \sim \text{Poisson}(\mu)$,

\[
P(X=k) = e^{-\mu} \frac{\mu^k}{k!}, \quad k = 0, 1, 2, \ldots
\]

\[
E(X) = \mu
\]

\[
SD(X) = \sqrt{\mu}
\]

Review of Poisson Dist.

For small $p$, $\text{Bin}(n,p) \approx \text{Pois}(np)$

If $N_1 \sim \text{Poisson}(\mu_1)$ independent

$N_2 \sim \text{Poisson}(\mu_2)$

then $N_1 + N_2 \sim \text{Poisson}(\mu_1 + \mu_2)$

Given: $N_1 + N_2 = 17$,

$N_1 \sim \text{Binomial}(n=17, p=\frac{\mu_1}{\mu_1+\mu_2})$

$X_1 + X_2 + \ldots + X_n \sim \text{Poisson}(\mu = \mu_1 + \mu_2)$

Indep. Bernoulli($p_1$), $p_1$ small

\[
\begin{array}{c|c|c|c|c|c|c}
 k & 0 & 1 & 2 & 3 & 4 & 5+
\hline
\text{actual} & 2.9 & 2.11 & 9.2 & 3.5 & 7 & 1
\hline
\text{expected} & 2.24 & 2.11 & 9.85 & 30.4 & 7.1 & 1.6
\hline
\text{RS.}(\mu, \mu_2) & 57.9 & 2.1 & 94.8 & 30.8 & 7.1 & 1.6
\end{array}
\]

Where

576 squares ($\frac{1}{4} \text{km}^2 @$)

537 total hits

\(\text{Poisson Scatter with?} \)

\(\mu \approx \frac{537}{576} = .932 \text{ per square} \)
Combining and Thinning Poisson Scatters

1. If you combine a Poisson scatter with intensity \( \lambda_1 \) with another Poisson scatter with intensity \( \lambda_2 \), then the total scatters are independent (if \( \lambda_1 \) and \( \lambda_2 \) are small).

2. If keep each point in a Poisson \( \lambda = 1 \), then scatter with prob \( x = \eta \) then you get a Poisson scatter with intensity \( \lambda \). Would be independent for \( \lambda \) Poissons.


Ex. 2 (Missprint) cont.

\[
P(k = 5 | N = 5) = \frac{e^{-5} 5^5}{5!} \cdot \frac{1}{3} = 0.0282
\]

Explanation of in

Poison scatter is independent

Not sure how to use it


Ex. 1

If Poisson intensity \( \lambda = 5 \), then expected number of hits per second are

\[
\lambda = 5 \times 60 \times 60 = 1800
\]

Mistake: 2. If square foot, scatters far away 

From \( \lambda = 20 \), over 10 seconds at 5 square foot.

2. Ifousse with \( \lambda = 1 \), then scatter with prob \( x = \eta \) then you get a Poisson scatter with intensity \( \lambda \). Would be independent for \( \lambda \) Poissons.
The Cancer-Cluster Myth

Alan Bender

"The reality is that [residential cancer-cluster] investigations are expensive, time-consuming, and complex."

"New Yorker," Feb. 8, 1999

Example

\[ P(N = 5 | M = 3) = \frac{\binom{5}{3} (0.01)^3 (0.99)^2}{\text{(sum of all probabilities)}} \]

E2: Michigan, continued again

E2: Michigan, continued again

Metland, California,

2,900 children, with cancer

Pesticides in groundwater?

"But that's not new."

L90

Low of large samples will resemble the whole population. Large random samples will resemble the whole population. Small random samples should resemble the whole population.

Not understanding how variable small random samples are.
Belief in the Law of Small Numbers, cont.:
- Tendency to see patterns in random data and to think they are due to something other than randomness.
Examples:
- Residential cancer-clusters
- Hot and cold streaks in B-ball
- Runs of good and bad luck in gambling

Cancer cluster references:
http://bear.cba.ufl.edu
/brnner/qmb3250/clusters.html
(Article by Gina Kolata)
http://www.dartmouth.edu
/~chance/chance_news/recent_news/
chance_news_99-02.html
(Jan-Feb 1999, Chance News)

Problem of Multiple Testing
If you investigate 100 different types of cancer in your area, probably at least a couple of types will be found to occur at an unusually high rate.
Texas-sharpsnoot Fallacy
Defining the population base around the observed cases.


Last time (Session 17)
Poisson scatters, cancer clusters
This time (Session 18)
Review of Chapters 1, 2, 3
Multinomial distribution
2. rev. 18
random counts versus random proportions
Proof God exists
Chapter 1
Outcome space $\Omega$; events $A \subseteq \Omega$.
Probability as limiting relative frequency.
Empirical Law of Averages says LRFs exist.
100 coin flips, $2^{100}$ outcomes
$\approx 95\%$ have 50±10 heads.
10,000 coin flips, $2^{10,000}$ outcomes
$\approx 95\%$ have 5000±100 heads.

Rules of Prob and Proportion

*equally likely 18.2

Chapter 2
Binomial distribution
Story
Probability Formula
Sums of indep Binomials $\binom{n}{k}$
mean and SD
Normal Approx.
Poisson Approx.

Hypergeometric Dist.
Story
Probabilities
(by counting comb.)
mean 18.5

Multinomial Distribution R155
Die with 3 red, 2 blue, 1 green face $\left(R_1 = \frac{1}{3}, R_2 = \frac{1}{3}, R_3 = \frac{1}{3}\right)$
$R = \#red, B = \#blue, G = \#green in 10 rolls$
$P(R=5, B=3, G=2)$

= $\binom{10}{5} \cdot \frac{1}{3}^5 \cdot \frac{2}{3}^3 \cdot \frac{1}{3}^2$

= 336

Independence

18.3
For \( n \) rolls, and
\( n_1 \geq D, n_2 \geq D, n_3 \geq D \) integers, \( n_1+n_2+n_3 \leq n \)

\[ P(n_1, n_2, n_3 | R = n_1, B = n_2, G = n_3) = \frac{n!}{n_1! n_2! n_3!} \frac{n_1}{P_r} \frac{n_2}{P_b} \frac{n_3}{P_g} \]

Since there are \( \frac{n!}{n_1! n_2! n_3!} \) arrangements
of \( n_1 \) 'r's, \( n_2 \) 'b's, \( n_3 \) 'g's,
and each arrangement has
probability \( \frac{n_1}{P_r} \frac{n_2}{P_b} \frac{n_3}{P_g} \)

(b) 3 of one kind, 4 of another

2 rev. 18 Roll 7 dice. Prob. of:
(a) exactly 3 sixes

(c) 2 fours, 2 fives, 3 sixes
Chapter 3 Random variables

Distributions
  joint, marginal, conditional
  independence

  Expectation (= weighted avg. of values)
  = long-run avg. (cf. CLT)
  linearity
  $E(XY) = E(X)E(Y)$ if $X, Y$ indep.

  Var and SD
  scaling and shifting
  additivity rule for var

(e) each number appears

Assume $P(\text{Boy}) = \frac{1}{2}$ for each birth.
  Births independent.

City A: 100 births per year
City B: 400 births per year

$N_A = \# \text{Boys in A}$

$P_A = \frac{N_A}{100} = \text{proportion of Boys in A}$

(e) sum is 9 or more
Which is bigger?
$P(N_A - 50 \geq 10)$ vs $P(N_B - 200 \geq 10)$

Actually, $P(\text{Boy}) = 0.513$

Basis of Arbuthnott’s 1710 proof of existence of God!

Data: 82 years of London birth records
Every year: More Boys than Girls
Assumption: If no God, then
$P(\text{More Boys}) = P(\text{More Girls}) \approx \frac{1}{2}$

So, data wildly improbable ($P = (\frac{1}{2})^{82}$)
If no God.

Inequalities: Markov & Chebyshev
CLT

Geometric series
Geometric r.v.’s
Story, prob. formula, memoryless prop.
Negative binomial r.v.’s
Story, prob. formula
Poisson r.v.’s
Sums
Scatter Theorem
Session 2.0

Continuous random variables densities, c.d.f.s, expectation infinitesimals

Def: A r.v. $X$ is continuous if it has probability 0 of equaling any fixed number $a$, i.e., $P(X=a)=0$, a.e. in $\mathbb{R}$.

Examples

1. $X \sim \text{Normal}(\mu, \sigma^2)$, which is std. normal on $z$-scale, i.e.,
   
   $Z = \frac{X-\mu}{\sigma} \sim \text{Normal}(0, 1)$.
Def. The distribution of r.v. $X$ is the assignment of probabilities $P\{X \in B\}$ to sets $B \subseteq \mathbb{R}$.

Remark: For "discrete" $X$ (as in Chap 2), enough to give possible values $x_1, x_2, x_3, \ldots$ and probabilities $P\{X = x_i\}$. Then for any $B \subseteq \mathbb{R}$,

$$P\{X \in B\} = \sum_{x_i \in B} P\{X = x_i\}$$

This won't work for continuous $X$. Why?

Fact: For any r.v. $X$ (discrete, or continuous, or "mixed"), distribution is determined by probs of intervals:

$$P\{X \in (a, b]\} = P\{a < X \leq b\}, \text{ all } a < b$$

(Example: $P\{X = b\} = \lim_{a \uparrow b} P\{X \in (a, b]\}$.)

Interval probs in turn, determined by probs $P\{X = b\} = \frac{dF}{dx}(b)$.

since

$$P[a < X \leq b] = P[X \leq b] - P[X \leq a]$$

$$= F(b) - F(a)$$

Def. $X$ has density $f$ if

$$P\{a < X \leq b\} = \int_a^b f(x) \, dx, \text{ all } a < b.$$

i.e., the probability that $X$ lands in interval $(a, b]$ equals area under curve $f$ over $(a, b]$.

Remark: All continuous r.v.'s in Stat 516 will have densities.

(Not true in general.)

Example: If $Z \sim N(0, 1)$, then $Z$ has density

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

and

$$\Phi(z) = \int_{-\infty}^z \phi(t) \, dt .$$

$$P\{0 < Z \leq 1\} = \int_0^1 \phi(z) \, dz$$

$$= \Phi(1) - \Phi(0) \text{ \text{ \text{ \text{ \text{ from Appendix}}}}$$

$$\approx 0.3413 - 0.5 = -0.1587$$
Sketch of $\Phi(z) = P(Z \leq z) = \text{std. normal cdf}$

\begin{align*}
\Phi(z) &= \int_{-\infty}^{z} \phi(t) \, dt \\
\Phi(z) &= \frac{1}{\sqrt{2\pi}} e^{-z^2/2}
\end{align*}

(2.1)

$W$ = random angle (in radians) values $[0, 2\pi]$

\[ cdf \\
F(w) = \begin{cases} 
0 & \text{if } b < 0 \\
\frac{b}{2\pi} & \text{if } 0 \leq b < 2\pi \\
1 & \text{if } b \geq 2\pi
\end{cases} \\
\]

Density $f(w) = \begin{cases} 
\frac{1}{2\pi} & \text{if } 0 < w < 2\pi \\
0 & \text{else}
\end{cases}$

(2.2)

Since $\Phi(z) = \int_{-\infty}^{z} \phi(t) \, dt$,,

slope of $\Phi(z)$ = height of $\phi(z)$

ie. $\frac{d}{dz} \Phi(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$

(Fundamental Theorem of Calculus)

(2.3)

Change of scale

Let $U = \frac{W}{2\pi} = \text{fraction of circle covered by angle } W$

\[ cdf \\
F(u) = \begin{cases} 
0 & \text{if } b < 0 \\
\frac{b}{2\pi} & \text{if } 0 \leq b < 2\pi \\
1 & \text{if } b \geq 2\pi
\end{cases} \\
\]

Density $f(u) = \begin{cases} 
\frac{1}{2\pi u} & \text{if } 0 < u < 1 \\
0 & \text{else}
\end{cases}$

(2.4)
Infinities

\[ P\left(x \leq x + dx\right) = P\left(x \leq dx\right) = f(x) \, dx \]

**Sum of probabilities:** \[ \sum f(x) \, dx = 1 \]

**Average of possible values, weighted by probabilities:**

\[ EX = \sum x \cdot f(x) \, dx \]

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

\[ f(x) = \frac{x^2}{\pi} \]

\[ \mathbb{E}(X) = \frac{\pi}{12} = 1.18 \]

\[ \text{Concentration of prob. near } x \]

\[ H \times \text{Concentration of prob. near } x \]

Integrals are sums over tiny pieces \( dx \).

\[ f(x) \, dx = \text{prob. of tiny piece } \, dx \]

\[ f(x) = \text{concentration of probability (per unit length) at } x \]

\[ \neq \text{a probability.} \]

---

\[ \Sigma \mathbb{E}(X) = \int_{-\infty}^{\infty} u \cdot f(u) \, du = \int_{-\infty}^{\infty} u \cdot \frac{1}{\pi} \, du = \frac{1}{2} \]

\[ \mathbb{E}(X^2) = \int_{-\infty}^{\infty} u^2 \cdot f(u) \, du = \int_{-\infty}^{\infty} u^2 \cdot \frac{1}{\pi} \, du = \frac{1}{3} \]

\[ \text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12} \]

Since \( W = 2\pi \cdot U \),

\[ \mathbb{E}(W) = 2\pi \cdot \mathbb{E}(U) = \pi \]

\[ \text{Var}(W) = \left(2\pi\right)^2 \cdot \text{Var}(U) = \frac{\pi^2}{9} \]

---

**Last time (Session 2.0)**

Continuous r.v.'s

**This time (Session 2.1)**

More on

"random point in a circle"

Example
Exercise 2. (cont.)

Find the probability density function of \( R \), the distance to the center, given by:

\[
f(r) = \begin{cases} 
2r & \text{if } 0 < r < 1 \\
0 & \text{else}
\end{cases}
\]

Formula for \( \Pr[R \leq b] = \frac{\text{area of circle of radius } b}{\text{area of unit circle}} \) for \( 0 \leq b \leq 1 \):

\[
\Pr[R \leq b] = \int_0^b 2r \, dr = r^2 \bigg|_0^b = b^2 \frac{\pi}{\pi} = \frac{b^2}{\pi}
\]

\[
\Pr[R > b] = 1 - \Pr[R \leq b] = 1 - \frac{b^2}{\pi}
\]

\[
\Pr[R = b] = \lim_{r \to b^-} \Pr[R \leq b] - \Pr[R \leq b^-] = 0
\]

1. For every \( b \in \mathbb{R} \),

\[
\Pr[R \leq b] = \text{area under } f_R \text{ to the left of } b
\]

2. For every \( a < b \),

\[
\Pr[a < R \leq b] = \text{area under } f_R \text{ over } (a, b]
\]

\[
= \int_a^b f_R(b) \, db = \Pr[R \leq b] - \Pr[R \leq a]
\]
Compare (1) and (2) on previous page with relationship between std. normal density \( \Phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \) and std. normal cdf \( \Phi \) (Appendix 5).

(3) For every \( r \) (except \( r = 0, r = b \))

\[
f(r) = \text{slope of } f_R \text{ at } r = \frac{f_R(r)}{dr}\Big|_{r\to b}.
\]

Since,

\[
F_R(r + dr) - F_R(r) = \Phi(r + dr) - \Phi(r) = \int_r^{r+dr} f(r) \, dr
\]

Now divide by \( dr \).

Median of \( R \) is (any) number \( m \) with \( P\{R < m\} = \frac{1}{2} \leq P\{R \leq m\} \).

\[m^2 = \frac{1}{2}, \quad m = \sqrt{\frac{1}{2}} = 0.707.\]

(1) Half of area under \( f \) is to left of \( m \), half is to right of \( m \).

(2) \( m \) solves \( F_R(m) = \frac{1}{2} \).

So \( m \) is where \( f_R \) crosses level \( \frac{1}{2} \).

\[\text{area of strip with height } 2\sqrt{1-x^2}, \text{ width } dx
\]

\[
\int_0^{\sqrt{2}/2} 2\sqrt{1-x^2} \, dx
\]

So:

\[
f_R(x) = \begin{cases} 2\sqrt{1-x^2}, & -1 < x < 1 \\ 0, & \text{elsewhere} \end{cases}
\]
Recall $U \sim U[0,1] \implies \text{var}(U) = \frac{1}{12}$

Note: $2U \sim U[0,2]$ and $\text{var}(2U) = 2^2 \times \text{var}(U) = \frac{4}{12} = \frac{1}{3}$

Q: Is the variance of $X$ above $\frac{1}{3}$? $\frac{1}{3}$

Are $X$ and $Y$ independent?

But $E(XY) = E(X)E(Y)$

Are $W$ and $R$ independent?

Independence of r.v.'s

General Def'n: $X$ and $Y$ independent if $P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$

for all sets of numbers $A, B$.

Recall: For $X, Y$ discrete, used $P\{X=x, Y=y\} = P\{X=x\}P\{Y=y\}$, all $x, y$

Still true that $E(XY) = E(X)E(Y)$

$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$