STAT 244
3/8/16
Lecture 18

Rank Order Statistics

There exist situations where the CLT does not hold, and the theory we have covered this quarter cannot be applied. There are still useful results, although they are never as sharp as the methods we learned this quarter.

Ex Recall Chebyshev's inequality:

$$Pr(|X-\mu| > k\sigma) \leq \frac{1}{k^2}$$

compare with normal dist.

Standard deviations from mean	$Pr(\lvert X-\mu\rvert > k\sigma)$ Chebyshev inequality upper bound	Normal Dist
1	1	~0.32
2	0.75	~0.046
3	~0.11	~0.0026

In these situations, it is often useful to sort the observations into rank order, and work with the ranks.

Ex. Y_i, $0 < i \le 5$ (actual example from J Chem Phys 148:646)

unsorted	rank order
$Y_1 = 8.6$	$X_1 = 28,549$
$Y_2 = 7.1$	$X_2 = 8.6$
$Y_3 = 6.9$	$X_3 = 7.4$
$Y_4 = 28,549.9$	$X_4 = 7.1$
$Y_5 = 7.4$	$X_5 = 6.9$

$$Y \xrightarrow{\text{sort}} X$$

$$\bar{X} = 5,715.8$$

sample median $= 7.4$

> Note: in this lecture only X_i denotes ordered data, Y_i the original un-data

remove X_1 (after finding independent reasons to do so). Then

$$\bar{X} = 7.5$$

sample median $= 7.25$

This is a rather extreme practical example of how the median is more robust than the mean.

Confidence Intervals for the Median

We want to estimate the population median, η, from ordered observations

$X_1, \ldots, X_n$, $1 \leq i \leq n$, $X_i \leq X_{i+1}$.

X iid from <u>some</u> cont. dist.

The definition of η is that

$$P(X_i > \eta) = P(X_i < \eta) = 0.5.$$

We'll want a C.I. in the form

(X_k, X_{n-k+1}), $k \leq \frac{n+1}{2}$, bigger k gives tighter interval.

Then

$$P(\eta > X_{n-k+1}) = \sum_{j=0}^{k-1} P(j \text{ observations} > \eta)$$

$$P(\eta < X_k) = \sum_{j=0}^{k-1} P(j \text{ observations} < \eta)$$

Because the X_i's are indep and identically distributed, $P(X_i > \eta)$ is a Bernoulli r.v. with $\theta = \frac{1}{2}$, and so

$$P(\text{exactly } j \text{ observations} > \eta) = \frac{1}{2^n}\binom{n}{j}$$

Hence

$$P(\eta > X_{n-k+1}) = \frac{1}{2^n} \sum_{j=0}^{k-1} \binom{n}{j}$$

and the

$$P(\eta < X_k) = \frac{1}{2^n} \sum_{j=0}^{k-1} \binom{n}{j}$$

so the coverage probability of (X_k, X_{n-k+1}) is

$$1 - \frac{1}{2^{n-1}} \sum_{j=0}^{k-1} \binom{n}{j}$$

This can be found from tables of the cumulative binomial distribution, since

$$P(Z \leq k-1) = \frac{1}{2^n} \sum_{j=0}^{k-1} \binom{n}{j}.$$

For a particular example with $n=26$ (Rice, p. 3[illegible]6) we have

K	$P(Y \leq K)$
5	.0012
6	.0047
7	.0145
8	.0378
9	.0843

choosing $k=8$, $P(Y<k) = 0.0145$.

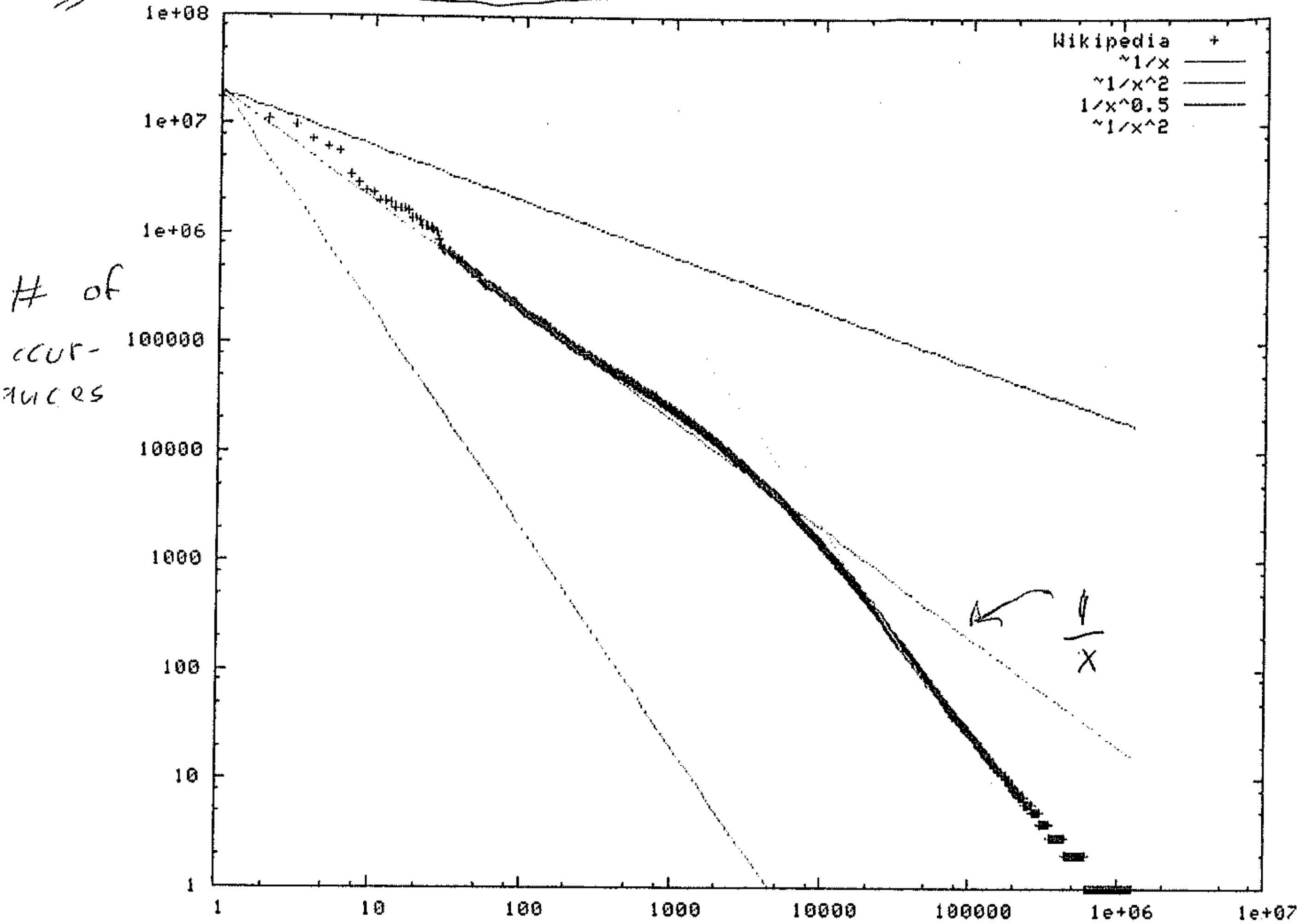
$P(Y<k) = P(Y > n-k+1)$

so $P(Y<8) = P(Y>19)$

and since $2 \cdot (0.0145) \approx 0.03$

(X_8, X_{19}) is a 97% CI.

Ex Rank-based distributions



Word frequency in Wikipedia, 11/26/06.

Rank data has distributions, too. Word frequeny rank in English is (largely) governed by Zipf's law, which says that frequency f, is distributed according to rank

$$F \sim \frac{1}{k}$$

Rank Distributions

The Zipf dist was a function of rank; let us now return to the distributions of the ranks themselves.

Ex Suppose Y_1, Y_2, Y_3 are 3 indep observations of $Y \sim f(y)$, $-\infty < y < \infty$. a continuous random variable.

Let us transform them to ordered X's by the function $\mathcal{F}(Y)$ below;

$$\mathcal{F}(y) = \begin{cases} X_1 = Y_1;\ Y_2 = X_2;\ Y_3 = X_3 & \text{if } Y_1 < Y_2 < Y_3 \\ X_1 = Y_1;\ X_2 = Y_3;\ X_3 = Y_2 & \text{if } Y_1 < Y_3 < Y_2 \\ X_1 = Y_2;\ X_2 = Y_1;\ X_3 = Y_3 & \text{if } Y_2 < Y_1 < Y_3 \\ X_1 = Y_2;\ X_2 = Y_3;\ X_3 = Y_1 & \text{if } Y_2 < Y_3 < Y_1 \\ X_1 = Y_3;\ X_2 = Y_1;\ X_3 = Y_2 & \text{if } Y_3 < Y_1 < Y_2 \\ X_1 = Y_3;\ X_2 = Y_2;\ X_3 = Y_1 & \text{if } Y_3 < Y_2 < Y_1 \end{cases}$$

Let's find the joint density of X_1, X_2, X_3.

The joint density of $\vec{Y}$ is

$$g(y_1, y_2, y_3) = f(y_1)\, f(y_2)\, f(y_3)$$

$$-\infty < y_i < \infty,\ i = 1, 2, 3$$

[note that The union of the six disjoint regions above is the 3D space with $-\infty < y_i < \infty$, $i = 1, 2, 3$]

Now we have $\tilde{X} = \tilde{r}(\tilde{Y})$, a function of a random variable, so the distribution of $\tilde{X}$, $h(x_1, x_2, x_3)$, is given by

$$h(\tilde{x}) = g(\tilde{r}^{-1}(y)) \left\| \frac{\partial \tilde{r}}{\partial y_k} \right\|$$

→ Jacobian; = 1 here.

So for the region $y_1 < y_2 < y_3$,

$$h(x_1, x_2, x_3) = f(y_1) f(y_2) f(y_3)$$

for the region $y_1 < y_3 < y_2$

$$h(x_1, x_2, x_3) = f(y_1) f(y_3) f(y_2)$$

... and similarly for the other 4 regions

⋮

$$h(x_1, x_2, x_3) = \underbrace{f(y_3) f(y_2) f(y_1)}$$

these are the same, and there are 6 pre images, so we add them to get

$$h(x_1, x_2, x_3) = 6 f(x_1) f(x_2) f(x_3),$$
$$-\infty < x_1 < x_2 < x_3 < \infty$$

or,

$$h(x_1, x_2, x_3) = 3!\, f(x_1) f(x_2) f(x_3)$$
$$-\infty < x_1 < x_2 < x_3 < \infty.$$

This generalizes to

$$h(x_1, \ldots, x_n) = n!\, f(x_1) \ldots f(x_n)$$
$$-\infty < x_1 < \cdots < x_n < \infty$$

It follows from this that the distribution of the area under the density function between any two ordered observations is independent of the form of the density function.

In other words, we are interested in permutations among the ranks.

The Wilcoxon Rank Test
(aka Mann-Whitney test)

Suppose there is a treatment group and a control group, with $m+n$ experimental replicates. We assign n units randomly to the control and m to the treatment.

Then:

1. Group all $m+n$ observations and rank them.
2. Sum the ranks from the control group and reject if the sum is too small or too large

Ex

	Treatment	Control
value →	1 (1) ← rank	6 (4)
	3 (2)	4 (3)
sum of ranks R	3	7

significant?

To construct the null hypothesis, we compute the distribution of R assuming assignment of each rank is equally likely.

In this example, there are $4! = 24$ such assignments, and $\binom{4}{2} = 6$ assignments to the control group are equally likely:

Ranks	R
$\{1, 2\}$	3
$\{1, 3\}$	4
$\{1, 4\}$	5
$\{2, 3\}$	5
$\{2, 4\}$	6
$\{3, 4\}$	7

Hence, under H_0

r	3	4	5	6	7
$P(R=r)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

So for the example, $P(R) = \frac{1}{6}$

In general, let T_x denote the sum of the ranks of $X_1, \ldots, X_m$. Using results covered in 245, we can show that under the null hypothesis,

$$E(T_X) = \frac{m(m+n+1)}{2}$$

$$Var(T_Y) = \frac{mn(m+n+1)}{12}$$

Moreover, if m and n are larger than about 7, T is well approximated by a normal distribution, and useful tables can be constructed.

Tune in next quarter with Prof. Wu for more...

STAT 244
WiQ Topic Summary

Topics

1-2: Permutations and combinations
Conditional Probability
Random Variables + Distributions (discrete)
- Bernoulli
- Binomial
- Negative Binomial
- Poisson

3. Continuous Random Variables
Functions of Random Variables
- Discrete
- Continuous
  - $Y = X^2$, $X \sim N, \Rightarrow Y \sim \chi^2_1$
  - $Y = aX + b \rightarrow$ std normal

4. Describing Distributions
- pmf / cdf
- median
- Expectations
  - expectation of function of r.v.
  - mean
  - linear transformations thereof
  - variance
  - linear transformations thereof

Multivariate Distributions
- Discrete
- Continuous
- Marginal Dists
- Conditional Dists

5. Expectations of Joint Distributions
   Expectations of Marginal Dists
   Covariance
  Law of Large Numbers
  Moments and Moment Generating Functions
  CLT derived (not proved) with MGFs

6. INFERENCE
  Bayesian Inference
  Bayes' Thm
   Discrete Dists
   Continuous Dists
   Mixed
  Beta Distribution
   Gamma Function
  Meaning of the Prior

7. Expectation of Beta Dist
  Bayesian Inference on Conjugate Priors
   Beta
   Normal
   Weighted average of prior + posterior

8. Likelihood
  Point estimation: $\hat{\theta}(x)$ from $f(x/\theta)$
  $MSE = Var + (Bias)^2$
  Maximum Likelihood $L(\theta) = f(x/\theta)$
   Find by setting $\frac{d}{d\theta} \log L(\theta) = 0$
   Check " verifying $\frac{d^2}{d\theta^2} \log L(\theta) < 0$

9. Distributions of Sums
  χ^2 density function derivation

10. Fisher's Thm
   If MLE found by setting $\frac{d}{d\theta} \log L(\theta) = 0$,
   $\hat{\theta}$ normally dist.
   $$\frac{1}{\tau^2} = E\left[\frac{d}{d\theta} \log f(x/\theta)\right]^2 = -E\left[\frac{d^2}{d\theta^2} \log f(x/\theta)\right]$$

10. (cont) <u>Fisher's Thm</u> (cont)

Cramer-Rao inequality

Case where $L(\theta) = \prod_{i=1}^{n} f(x_i | \theta)$, $\tau_n = \frac{\tau^2}{n}$

Fisher <u>doesn't</u> apply if max at edge of domain

11. Fisher's Thm for multivariate case

Although MLE's are often biased, $\bar{X}$ unbiased for μ, s^2 for σ^2

Non-rigorous proof of Fisher's Thm

11A. Sufficient Statistics and Neymann Factorization Thm; Rao-Blackwell Thm

12. <u>Hypothesis</u> <u>Testing</u>

<u>Simple</u> <u>Hypotheses</u>: α, β, $\pi = 1 - \beta$

Likelihood Ratios, which are best 'cause

Neyman - Pearson Lemma

Power Functions

Uniformly Most Powerful Tests

13. Proof of Neyman-Pearson

<u>LR</u> $\lambda = \frac{\max_{\theta_0\text{'s}} L(\hat{\theta}_0)}{\max_{\text{all } \theta\text{'s}} L(\hat{\theta}_1)}$ Reject H_0 if $\lambda < \lambda_c$, $P(\lambda < \lambda_c | H_0) \leq \alpha$

The Multinomial Distribution

14. Composite Hypotheses

Reject H_0 if $\lambda < \lambda_c \Rightarrow -\log \lambda > C$

$\tilde{\chi}$ $-\log \lambda \cong \frac{(\text{observed} - \text{expected})^2}{\text{expected}} = \chi^2$, dist chi-square

Multinomial, k outcomes, χ^2_{k-1} (test 'fairness')

Multinomial, k outcomes, p $\hat{\theta}$'s by MLE, χ^2_{k-p-1} (test form of distribution)

Contingency Tables - $\begin{matrix} r-1 \\ c-1 \end{matrix}$

Depend on row and column totals,

Multinomial $(r-1)(c-1)$ d.f.

H_0: cell probs product of marginal probs

15. Tests of Homogeneity
Multinomial Xir trials
H_0: probs the same all rows
Same χ^2, same d.f. as other multinomial tests
χ^2 an approx.
all expected > 3.5 (?)
80% " > 5 (?)
etc.
No expected $= 0$!
opinion differs on exact borders

LAST HOMEWORK

P-values
smallest α to reject Ho

Confidence Intervals
random interval with $1-\alpha$ prob of containing θ

16. dual to hypothesis tests

Questions about these topics easier

More χ^2:
Multinomial χ^2 for Poisson
Small χ^2 and Mendel's peas
Meta-analysis by Fisher's method of combination

16. Rank Order Statistics
Hypothesis testing for the median.