Probability Theory:

In random experiments we are interested in the occurance of events represented by sets.

General idea: Probability is about measuring the relative sizes of sets.

Basic structure:

Suppose an experiment yields a random outcome:

(1) Outcomes are element of the sample space (Ω) which is the set of all possible outcomes.

(2) Events are a special class of subsets of the sample space (called "measurable sets").

(3) The probability of an event is the size of the set relative to the whole sample space (Ω).

Set Theory:

- A set is a collection of elements
- A is subset of set B if every element of A is also an element of B
  $x \in A \rightarrow x \in B$.
- The universal set (U or X) is a set that contains all possible objects of interest.
- The empty set is a set with no elements ($\emptyset$).
- The union of two sets $A \cup B = \{x : x \in A \text{ OR } x \in B\}$.
- The intersection of two sets $A \cap B = \{x : x \in A \text{ AND } x \in B\}$.
  Two sets are disjoint if $A \cap B = \emptyset$.
- The complement of a set $A^c = \{x : x \notin A\}$.

- The difference of two sets $A - B = \{x : x \in A \text{ AND } x \notin B\}$.
- A function ($f: X \to Y$) is a mapping between sets such that eache element of the Domain $x \in X$ maps to a unique element in the range $f(x) \in Y$.
- The powerset of a set X is the set that contains all subsets of X:

$$2^X = \{A : A \subset X\}.$$

- A class is a set-of-sets.
- A class is a σ-algebra ($\mathcal{Q} \subset 2^X$) iff $\mathcal{Q}$ is C-U-T.
  - C: closed under complement. $A \in \mathcal{Q} \to A^c \in \mathcal{Q}$.
  - U: closed under countable unions $A_1, A_2 \ldots A_n \in \mathcal{Q} \to \bigcup_{k=1}^{n} A_k \in \mathcal{Q}$
  - T: total set in $\mathcal{Q}$. $X \in \mathcal{Q}$

The axioms of probability:

Suppose E is a random experiment with sample space X and with σ-algebra $\mathcal{Q}$. A probability measure for E assigns a number called the probability of A to each event $A \in \mathcal{Q}$ that satisfies:

1. $P[A] \geq 0$.
2. $P[X] = 1$
3. If $A \cap B = \emptyset$ then $P[A \cup B] = P[A] + P[B]$.

(3') If $A_1, A_2, \ldots$ is a sequence of events with $A_i \cap A_j = \emptyset$ for $i \neq j$ then $P\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} P[A_k]$.

Random variables:

The use of random variables allows us to ignore the "fine" structure of Ω and focus on the probabity P.

- The cumulative distribution fuction (cdf) of the random variable X is defined for all real numbers: $F(x) = P(X \leq x)$.

- For a discrete random variable (finite or countable distinct values) its probability mass function (pmf): $p(x) = P\{X = x\}$.

shorthand for: $P(\{w : X(w) = x\}$

It follows by axioms: $\sum_{i=1}^{\infty} p(x_i) = 1$

- For continuous set of possible outcomes:

$$P\{X \in C\} = \int_C f(x)\,dx$$

where f is called the probability density fuction (pdf) it may not exist.

Notice: $F(a) = P\{X \in (-\infty, a)\} = \int_{-\infty}^{a} f(x)\,dx$.

differentiating gives: $\frac{d}{da} F(a) = f(a)$.

intuitively: $\varepsilon \cdot f(a) \approx \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x)\,dx = P\{a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2}\}$.

Expectation: If X is a discrete random variable.

$$E[x] = \sum_{i=1} x_i \cdot P[X = x_i].$$

a "weighted average" of possible values.

Continuous random variable

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

If X is a discrete random variable with p.m.f. $p(x)$.

then $$E[g(x)] = \sum_x g(x) \cdot p(x).$$

pdf $f(x)$:

$$E[g(x)] = \int_x g(x) f(x)\, dx$$

Variance: If X is a random variable with mean μ. then

the variance of X, $$Var[X] = E[(X-\mu)^2].$$

$$Var[x] = E[x^2] - (E[x])^2.$$

Standard deviation: $\sigma_x = \sqrt{Var[X]}$ μ_x and σ_x lie on the same axis.

Covariance: The covariance of two random variables X and Y,

$$Cov(X, Y) = E[(X-\mu_x)(Y-\mu_Y)].$$

where $\mu_x = E[X]$ and $\mu_Y = E[Y]$.

$$Cov(X,Y) = E[X \cdot Y] - E[X] \cdot E[Y].$$

If two random variables X and Y are independent then

$$Cov[X, Y] = 0.$$

***
Note: If $Cov(X, Y) = 0$ then the random variables are uncorrelated.
Not necesarily independent.

EE511 17 Jan 2017

Common propability distributions (pdfs).

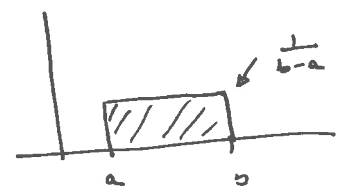
- B binomial (bernoulli and related) $\longrightarrow$
- E exponential
- G gaussian
- C cauchy
- U uniform
- P poisson.

		Sampling	
		w. Replacement	w/o. replacement
Outcomes	2	Binomial Negative Binomial -geometric	Hypergeometric
	≥ 3	multinomial	multivariate hypergeom.

Today:

Uniform: over interval $[a,b]$ (usual "standard" $[a,b]=[0,1]$).

$$f(x)=\begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{else} \end{cases}$$



$$E[x]= \frac{b+a}{2}.$$

$$V[x] \quad \left(E[x^2]= \frac{a^2+b^2+ab}{3} \right) = \frac{1}{12}(b-a)^2$$

matlab: rand( )

Q (home): generate a uniform random sample: $X \sim U[-2,3]$.

Statistics Concepts:

Data controls: let data tell its own story.

data trumps math (hypotheses).

Sample mean:

Average value of raw data:

$$\overline{X}_n = \frac{1}{n} \cdot \sum_{k=1}^{n} X_k.$$

Sample variance:

The "spread" of the data

$$S_x^2 = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \overline{X}_n)^2$$

Defn: $X_1, X_2 \ldots$ is a random sample iff $X_1, X_2, \ldots, X_n$ is i.i.d.

independent and identically dist.

Defn: T is a statistic iff $T = T(X_1, \ldots, X_n)$ for random samples $X_1, \ldots X_n$

Observe a random sample $X_1, X_2, \ldots, X_n$ to get a realization of that sample (a number)

$$\begin{array}{cccc} X_1 & X_2 & \ldots & X_n \\ \downarrow & \downarrow & & \downarrow \\ x_1 & x_2 & \ldots & x_n \end{array}$$

$\therefore$ $\overline{X}_n$ leaves one realization: $\overline{x}_n$

Confidence intervals

Idea: interval estimates are more reliable than point estimates.

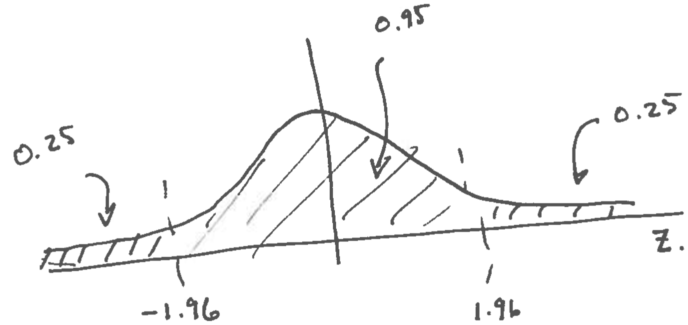
- Less data $\longrightarrow$ wider interval.
- point estimate $\pm$ margin of error (m.o.e.).

Ex: $\mu \pm \sigma$. $\longrightarrow$ $\bar{X}_n \pm$ m.o.e.

random variables.

Canonical example: 95% confidence interval for the population mean.

$$X_n \pm Z_{0.025} \frac{\sigma}{\sqrt{n}}.$$ (C.I. is a random set).



± 1.96 (≈ 2) $\longleftrightarrow$ 95% C.I. for μ.

Bootstrap C.I. for μ.

- no pdf assumptions.
- Sample <u>with replacement</u> to create "synthetic data"

Ex: Generate 100 random samples. $\rightarrow$
1. Form 1000 resamples of size 100 each.
2. For each resample compute sample mean.
3. rank computed sample means

$\bar{X}_{(1)} \le \bar{X}_{(2)} \cdots \le \bar{X}_{(1000)}$ then 95% C.I. $= (\bar{X}_{(25)}, \bar{X}_{(975)})$.

same process for σ^2 / σ / ...

χ^2 goodness of fit test.

$$\chi^2 = \sum_{j=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

O_i : observed

E_i : expected

idea: group data into "bins" and compare the number expected with the number observed.

$$= \sum_{j=1}^{K} \frac{(n_i - E[n_i])^2}{E[n_i]} \sim \chi^2(K-1)$$

degrees of freedom.

for large n. (By CLT, multinomial extension of proportions).

<u>Important:</u> "Bins" : 1. at least 5 samples/bin

2. $E[n_i] \geq 5 \quad \forall n$.

then test w/ p-value from χ^2 table. (use builtin functions if available)

Smaller $\chi^2 \longrightarrow$ better fit (model better explains the data).

<u>Ex:</u> H_0: $X \sim$ Uniform

H_A: $X \not\sim$ uniform

<u>Ex</u>: χ^2-test for Poisson pdf. Recall $p(x) = \frac{e^{-\lambda} \cdot \lambda^K}{K!}$

mean = variance = λ.

Lung cancer survival in a hospital. $n = 450$.

Construct a "1 way" table:

	Survival time (in years)				
Observed	0	1	2	3	≥ 4
# patients	60	110	125	88	67.

Suppose average survival time is $\lambda = 2.1$ years. $\therefore$ if poisson $\lambda = 2.1$ yrs.

<u>Conjecture</u>: $X \sim p(2.1)$.

$\therefore$ H_0: pdf = $P(2.1)$

H_A: pdf $\neq P(2.1)$ at α-significance level.

Given H_0: $\therefore P(x=0) = e^{-2.1} = .122 = P(0)$. $P(0) \cdot 450 = 54.9$ patients.

$P(x=1) = 2.1 \cdot e^{-2.1} = .257 = P(1)$ $P(1) \cdot 450 = 115.65$ patients.

$P(x=2) = \frac{(2.1)^2 \cdot e^{-2.1}}{2!} = .27 = P(2)$ $P(2) \cdot 450 = 121.5$

$P(x=3) = \frac{(2.1)^3 e^{-2.1}}{3!} = .189 = P(3$ $P(3) \cdot 450 = 85.01$.

P(4), multinomial structure so $P(4) = 1 - [P(0) + P[1] + P[2] + P[3]] = .189$.

$P(4) \cdot 450 = 72.9$ →

Ex (cont.)

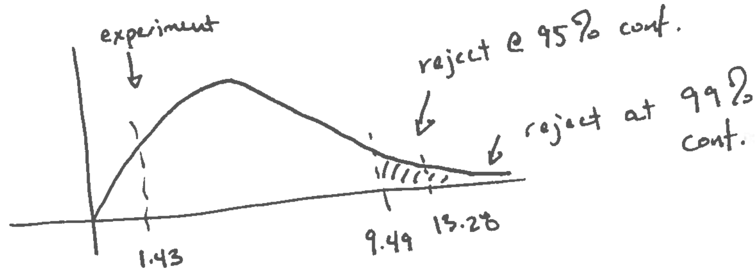
years	0	1	2	3	≥4
observed	60	110	125	88	67
expected	54.9	115.65	121.5	85.05	72.9

degrees of freedom $= k-1$
$= 5-1$
$= 4$ $=$ # categories $- 1$.

$$\therefore \chi^2 = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} = \cdots = 1.43 \quad \text{(small)}$$

Table: $\chi^2_{0.05}(4) = 9.49$

$\chi^2_{0.01}(4) = 13.28$

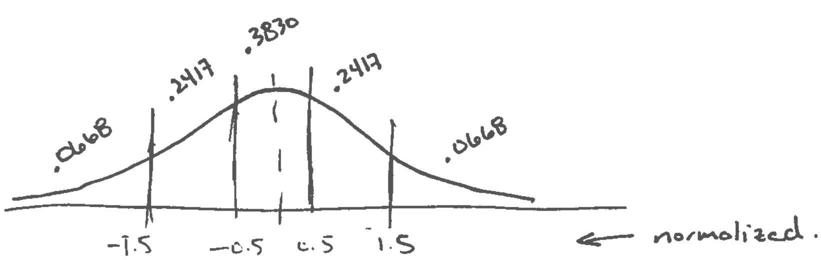


$1.43 \ll 9.49$.

$\therefore$ "Accept" H_0, $X \sim P(2.1)$.

Ex: For bell-curve hypothesis, $X \sim N(\mu, \sigma^2)$, Partition bell curve into categories.

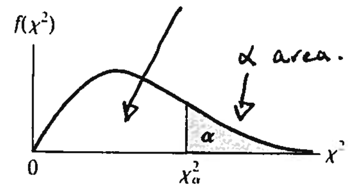
e.g. $k=5$



degrees of freedom: $k-1 = 5-1 = 4$.

TABLE 8 Critical Values of χ^2

confidence Level = 1 − α



Degrees of Freedom	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$
1	.0000393	.0001571	.0009821	.0039321	.0157908
2	.0100251	.0201007	.0506356	.102587	.210720
3	.0717212	.114832	.215795	.351846	.584375
4	.206990	.297110	.484419	.710721	1.063623
5	.411740	.554300	.831211	1.145476	1.61031
6	.675727	0.872085	1.237347	1.63539	2.20413
7	.989265	1.239043	1.68987	2.16735	2.83311
8	1.344419	1.646482	2.17973	2.73264	3.48954
9	1.734926	2.087912	2.70039	3.32511	4.16816
10	2.15585	2.55821	3.24697	3.94030	4.86518
11	2.60321	3.05347	3.81575	4.57481	5.57779
12	3.07382	3.57056	4.40379	5.22603	6.30380
13	3.56503	4.10691	5.00874	5.89186	7.04150
14	4.07468	4.66043	5.62872	6.57063	7.78953
15	4.60094	5.22935	6.26214	7.26094	8.54675
16	5.14224	5.81221	6.90766	7.96164	9.31223
17	5.69724	6.40776	7.56418	8.67176	10.0852
18	6.26481	7.01491	8.23075	9.39046	10.8649
19	6.84398	7.63273	8.90655	10.1170	11.6509
20	7.43386	8.26040	9.59083	10.8508	12.4426
21	8.03366	8.89720	10.28293	11.5913	13.2396
22	8.64272	9.54249	10.9823	12.3380	14.0415
23	9.26042	10.19567	11.6885	13.0905	14.8479
24	9.88623	10.8564	12.4011	13.8484	15.6587
25	10.5197	11.5240	13.1197	14.6114	16.4734
26	11.1603	12.1981	13.8439	15.3791	17.2919
27	11.8076	12.8786	14.5733	16.1513	18.1138
28	12.4613	13.5648	15.3079	16.9279	18.9392
29	13.1211	14.2565	16.0471	17.7083	19.7677
30	13.7867	14.9535	16.7908	18.4926	20.5992
40	20.7065	22.1643	24.4331	26.5093	29.0505
50	27.9907	29.7067	32.3574	34.7642	37.6886
60	35.5346	37.4848	40.4817	43.1879	46.4589
70	43.2752	45.4418	48.7576	51.7393	55.3290
80	51.1720	53.5400	57.1532	60.3915	64.2778
90	59.1963	61.7541	65.6466	69.1260	73.2912
100	67.3276	70.0648	74.2219	77.9295	82.3581

95% confidence level

99% confidence interval

4-degrees of freedom

Degrees of Freedom	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	2.70554	3.84146	5.02389	6.63490	7.87944
2	4.60517	5.99147	7.37776	9.21034	10.5966
3	6.25139	7.81473	9.34840	11.3449	12.8381
4	7.77944	$\boxed{9.48773}$	11.1433	$\boxed{13.2767}$	14.8602
5	9.23635	11.0705	12.8325	15.0863	16.7496
6	10.6446	12.5916	14.4494	16.8119	18.5476
7	12.0170	14.0671	16.0128	18.4753	20.2777
8	13.3616	15.5073	17.5346	20.0902	21.9550
9	14.6837	16.9190	19.0228	21.6660	23.5893
10	15.9871	18.3070	20.4831	23.2093	25.1882
11	17.2750	19.6751	21.9200	24.7250	26.7569
12	18.5494	21.0261	23.3367	26.2170	28.2995
13	19.8119	22.3621	24.7356	27.6883	29.8194
14	21.0642	23.6848	26.1190	29.1413	31.3193
15	22.3072	24.9958	27.4884	30.5779	32.8013
16	23.5418	26.2962	28.8454	31.9999	34.2672
17	24.7690	27.5871	30.1910	33.4087	35.7185
18	25.9894	28.8693	31.5264	34.8053	37.1564
19	27.2036	30.1435	32.8523	36.1908	38.5822
20	28.4120	31.4104	34.1696	37.5662	39.9968
21	29.6151	32.6705	35.4789	38.9321	41.4010
22	30.8133	33.9244	36.7807	40.2894	42.7956
23	32.0069	35.1725	38.0757	41.6384	44.1813
24	33.1963	36.4151	39.3641	42.9798	45.5585
25	34.3816	37.6525	40.6465	44.3141	46.9278
26	35.5631	38.8852	41.9232	45.6417	48.2899
27	36.7412	40.1133	43.1944	46.9630	49.6449
28	37.9159	41.3372	44.4607	48.2782	50.9933
29	39.0875	42.5569	45.7222	49.5879	52.3356
30	40.2560	43.7729	46.9792	50.8922	53.6720
40	51.8050	55.7585	59.3417	63.6907	66.7659
50	63.1671	67.5048	71.4202	76.1539	79.4900
60	74.3970	79.0819	83.2976	88.3794	91.9517
70	85.5271	90.5312	95.0231	100.425	104.215
80	96.5782	101.879	106.629	112.329	116.321
90	107.565	113.145	118.136	124.116	128.299
100	118.498	124.342	129.561	135.807	140.169

Source: From Thompson, C. M. "Tables of the percentage points of the χ^2-distribution." *Biometrika*, 1941, Vol. 32, pp. 188–189. Reproduced by permisson of the *Biometrika* Trustees.