Interpreting Results [Law82]

Descriptive Statistics

Experiment

Data

Sample statistics, like sample mean, sample variance, ...

Inferential Statistics

Model for probabilistic mechanism, like joint densities, ...

Estimates with confidence intervals

Inferences

Predictions
Suppose \( X \) is a random variable with known prob. density \( p(x) \).

**Mean**
\[
\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} x \ p(x) \, dx \quad (\text{discrete} \quad \sum x_i \ p_i)
\]

**Variance**
\[
\sigma^2 = \mathbb{E}[(X-\mu)^2] \\
= \mathbb{E}[x^2] - 2 \, \mathbb{E}[x \mu] + \mathbb{E}[\mu^2] \\
= \mathbb{E}[x^2] - \mu^2
\]
\( \sigma \) is called standard deviation

---

Distinguish from sample mean, sample variance:

Suppose we have \( N \) independent observations of \( X \), \( X_1, X_2, \ldots X_N \)

- **Sample mean** = \( \frac{1}{N} \sum_{i=1}^{N} x_i = \bar{x} \) (a "statistic")
  \[ \mathbb{E}[\bar{x}] = \mu \quad \text{unbiased} \]

- **Sample variance** = \( s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \) (a random variable)

Why divide by \( N-1 \)? \( \Rightarrow \mathbb{E}[s^2] = \sigma^2 \) unbiased
Intuitively, the $N$ differences $(x_i - \bar{x})$

are not independent, because $\sum_{i=1}^{N} (x_i - \bar{x}) = 0$

\[ \therefore \text{there are only } (N-1) \text{ degrees of freedom in statistic} \]

A proof is straightforward but requires some algebra

Some details, then:

Lemma Suppose samples $x_1, \ldots, x_N$ come from distribution with mean $\mu$ & variance $\sigma^2$, and are independent.

Then

\[ E(\bar{x} - \mu)^2 = \frac{\sigma^2}{N} \]

→ shows that standard deviation decreases as $1/N$

Proof algebra

Then

\[ E[s^2] = \frac{1}{N-1} E \left[ \sum_{i=1}^{N} (x_i - \bar{x})^2 \right] \]

\[ = \frac{1}{N-1} E \left[ \sum_{i=1}^{N} x_i^2 - N\bar{x}^2 \right] \]

\[ = \frac{1}{N-1} \left[ N(\sigma^2 + \mu^2) - N \left( \frac{\sigma^2}{N} + \mu^2 \right) \right] \]

\[ = \sigma^2 \quad \checkmark \]
A computational point:

\[ S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \]

Can compute by first finding \( \bar{x} \), then direct evaluation.

But

\[ S^2 = \frac{1}{N-1} \left[ \sum x_i^2 - 2 \bar{x} \sum x_i + \sum (\bar{x})^2 \right] \]

\[ = \frac{1}{N-1} \left[ \sum x_i^2 - 2 (\bar{x})^2 N + N (\bar{x})^2 \right] \]

\[ = \frac{1}{N-1} \left[ \sum x_i^2 - N (\bar{x})^2 \right] \]

\[ S^2 = \frac{\sum x_i^2 - N (\bar{x})^2}{N-1} \]

- fewer operations
- more accurate
**Importance of Gaussian (Normal) Distribution**

\[ p(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \]

often denoted \( N(\mu, \sigma^2) \). \( E(x) = \mu \) \( var(x) = \sigma^2 \)

We often deal with \( z = \frac{x - \mu}{\sigma} \) normalized Gaussian \( N(0, 1) \)

Tails are small

\[ \text{prob. } = 0.00135 \]

\[ \text{prob. of deviating more than } 3\sigma \text{ from mean } = 0.0027 \]

\[ \text{prob. of not } = 99.73\% \]

\[ \text{a prob. of deviating less than } 2\sigma = 95.45\% \]

**Why so important?**

Sum of independent observations converge to Gaussian under very general circumstances.

In nature, events that result from many small, independent effects tend to be Gaussian.
Central Limit Theorem

Suppose we sample $x_1, \ldots, x_n$ from a distribution with mean $\mu$ and variance $\sigma^2$.

Let $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ as usual.

Then

$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \rightarrow N(0,1)$

Normalized random variable

Notice this is true for any parent distribution. (See probability books for technical details & conditions.)

Important Properties of Normal Distribution

1. Linear combination of Normals is also Normal
2. Normal has maximum entropy for given $\sigma$.
3. Least-squares becomes maximum likelihood
4. Many derived random variables have analytically known densities
5. Sample mean and variance of $n$ identical, independent samples are independent; the sample mean is Normal $\bar{x}_n \sim N(\mu, \sigma/\sqrt{n})$. 
Summary of Distributions of random Variables derived from Normal

Let $X_i$ be $n$ independent, distributed samples from $N(\mu, \sigma)$.

**Sample Mean**

\[ \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{is distrib as } N(\mu, \sigma/\sqrt{n}) \]

**Sample Variance**

\[ S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \]

Then

\[ U = \frac{(n-1) S_n^2}{\sigma^2} \quad \text{(Normalized)} \]

has a $\chi^2$-distribution with $(n-1)$ degrees of freedom:

\[ p(\chi^2) = \left[ 2^{n/2} \Gamma(n/2) \right]^{-1} (\chi^2)^{n/2 - 1} e^{-\chi^2/2} \quad \chi^2 \geq 0 \]

\[ E[U] = n-1 \]

\[ \text{var}[U] = 2(n-1) \]

**Sample Mean Unknown Variance**

\[ \frac{\bar{X} - \mu}{S_n / \sqrt{n}} \quad \text{has a } t \text{-distribution} \]

with $(n-1)$ degrees of freedom:

\[ p(t) = \frac{\Gamma \left( \frac{n+1}{2} \right)}{\Gamma \left( \frac{n}{2} \right)} \frac{1}{\sqrt{n}} \frac{1}{\left[ 1 + \frac{t^2}{n} \right]^{\frac{n+1}{2}}} \]

"Student- $t$" = W.S. Gosset

broader tails than $N(\cdot)$
Confidence Intervals

Assume again \( x_1, \ldots, x_n \) i.i.d. Normal.
We want to know how far \( \bar{x}_n \) might be from \( \mu \).

We know \( \frac{\bar{x}_n - \mu}{s_n / \sqrt{n}} \) is Student-t distributed.

1. Pick \( q_1, q_2 \in \text{Student-t with (n-1) degrees of freedom has 99\% prob} \) of lying between \( q_1, q_2 \):

\[
\frac{\bar{x}_n - \mu}{s_n / \sqrt{n}} < q_2 \quad \text{and} \quad q_1 < \frac{\bar{x}_n - \mu}{s_n / \sqrt{n}}
\]

From tables

2. Prob. \( \left\{ q_1 < \frac{\bar{x}_n - \mu}{s_n / \sqrt{n}} < q_2 \right\} = 0.99 \)

\[
\Rightarrow \quad \text{prob.} \left\{ \bar{x}_n - \frac{q_2}{\sqrt{n}} s_n < \mu < \bar{x}_n - \frac{q_1}{\sqrt{n}} s_n \right\} = 0.99
\]

\[ \bar{x}_n \]

\[ -\frac{q_2}{\sqrt{n}} s_n \quad -\frac{q_1}{\sqrt{n}} s_n \]

"99\% confidence interval"

\( \text{prob. this line is 99\%} \)
Confidence Interval for $\sigma^2$

Again, assume $x_1, \ldots, x_n$ i.i.d. Normal.

We know $(n-1) \frac{S_n^2}{\sigma^2}$ is $\chi^2$-distributed with $n-1$ D.O.F.

1. Pick $g_1, g_2 \in \mathbb{R}$

   $$\text{Prob.} \left\{ \chi^2 \text{ n-1 D.O.F. between } g_1, g_2 \right\} = 0.99$$

2. $$\text{Prob.} \left\{ g_1 < \frac{(n-1)S_n^2}{\sigma^2} < g_2 \right\} = 0.99$$

   $$\Rightarrow \text{Prob.} \left\{ \frac{(n-1)S_n^2}{g_2} < \sigma^2 < \frac{(n-1)S_n^2}{g_1} \right\} = 0.99$$

   $$\sigma^2$$

   $$(n-1)S_n^2 \quad \quad (n-1)S_n^2$$

   $\frac{g_2}{g_1}$

   Prob. that true $\sigma$ is here is 99%.

CAVEAT: Must be close to Normal to be valid

$\rightarrow$ Can often invoke Central Limit theorem to justify.
Assume prob. of arrival in any interval proportional to \( \Delta t \), say \( \mu \Delta t \), where \( \mu = \text{avg. # of arrivals/sec} \).

- independent of all other arrivals in other intervals.
- small enough so we can neglect prob. of >1 arrivals.

Then

\[
\text{prob. } \{ \text{no arrivals in intervals } 1 \text{ to } n \} = (1 - \mu \Delta t)^n
\]

\[
= (1 - \mu \Delta t)^{t/\Delta t}
\]

Let \( \Delta t \to 0 \).

\[
\lim (1-x)^{1/x} = e^{-1}, \text{ so}
\]

\[
\text{prob. } \{ \text{no arrivals till } t \} = e^{-\mu t}
\]

\[
\text{prob. } \{ \text{first arrival before } t \} = 1 - e^{-\mu t}
\]

\[
\text{prob. density } \{ \text{time } t \text{ to first event} \}
\]

\[
= p(t) = \frac{d}{dt} [1 - e^{-\mu t}] = \mu e^{-\mu t}
\]

"Exponentially distributed" "Poisson arrivals"
two methods for generating such random arrivals:

I. take small \( \Delta t \), flip coin with probability \( \mu \) at each interval

II. generate \( x \) uniform on \([0,1]\),

use \( \left\lfloor -\ln(x) \right\rfloor \mu \) to determine next arrival

(see transformation method, p. 1.2.11)