6.1: Statistics And Their Distributions

Taylor

University of Virginia

Definitions

- Random variables are typically denoted with letters at the end of the alphabet (e.g. X,Y, etc.)
- Before we observe a sample, the data is random $(X_1, \ldots X_N)$
- The observations or data we observe are lower-case $(x_1, \ldots x_n)$
- for more info see Chapter 3

Definitions

- The same idea applies to functions of the rvs
- Example: $f(X_1, \ldots, X_n) = \frac{1}{n}\sum_i X_i$
- if we have the data already it's $f(x_1, \ldots, x_n) = \frac{1}{n}\sum_i x_i$
- This is what a **statistic** is: a quantity calculated from sample data.
- We're usually referring to the random variable (capital letters) because we like to talk about it's distribution/moments etc.

Examples

Some very ubiquitous examples:

- $f(X_1, \ldots, X_n) = \frac{1}{n} \sum_i X_i = \bar{X}$
- $f(X_1, \ldots, X_n) = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2 = S^2$
- $f(X_1, \ldots, X_n) = max(X_1, \ldots, X_n) = X_{(n)}$
- $f(X_1, \ldots, X_n) = min(X_1, \ldots, X_n) = X_{(1)}$

Definitions

- Every sample statistic has a probability distribution
- sometimes it's easy to calculate (e.g. the mean of a bunch of normal random variables)
- sometimes it's difficult or impossible
- The probability distribution of a statistic is often called a **sampling distribution**
- A note on notation: I'll usually denote general probability distributions with a p or an f

Definitions

- Many analyses assume that the data are being generated from a common probability distribution
- The most tractable situation is when each data point/rv comes from the same distribution, and no data point affects any other data point
- the book calls this a **random sample** (page 287)
- I will just say (i.i.d) which stands for *independent and identically distributed*
- I will denote it like this $\stackrel{iid}{\sim}$
- example: $X_1, \ldots, X_n \stackrel{iid}{\sim} \text{Normal}(\mu, \sigma^2)$

Example

Our first example (Example 6.3 on page 290):

Let $X_1, X_2 \overset{iid}{\sim}$ Exponential(λ). What is the distribution of $T_0 = \sum X_i$? First we'll find it's cumulative distribution function (cdf) $F_{T_0}(t)$; then we'll take the derivative to get the density $f_{T_0}(t)$. Hopefully it'll be a familiar distribution again...

Example

$$
\begin{aligned}
F_{T_0}(t) &= P(X_1 + X_2 \le t) \\
&= \iint_{\{(x_1,x_2):x_1+x_2\le t\}} p(x_1, x_2)\,dx_1\,dx_2 \\
&= \iint_{\{(x_1,x_2):x_1+x_2\le t\}} p(x_1)p(x_2)\,dx_1\,dx_2 \\
&= \int_0^t \int_0^{t-x_1} \lambda e^{-\lambda x_1} \lambda e^{-\lambda x_2}\,dx_2\,dx_1 \\
&= \int_0^t \lambda e^{-\lambda x_1} \left[\int_0^{t-x_1} \lambda e^{-\lambda x_2}\,dx_2\right] dx_1 \\
&= \int_0^t \lambda e^{-\lambda x_1} \left[-e^{-\lambda x_2}\right]_0^{t-x_1} dx_1 \\
&= \int_0^t \left(\lambda e^{-\lambda x_1} - \lambda e^{-\lambda t}\right) dx_1
\end{aligned}
$$

Example

$$
\begin{aligned}
&= \int_0^t \left(\lambda e^{-\lambda x_1} - \lambda e^{-\lambda t} \right) dx_1 \\
&= \int_0^t \lambda e^{-\lambda x_1} dx_1 - t\lambda e^{-\lambda t} \\
&= \left[-e^{-\lambda x_1} \right]_0^t - t\lambda e^{-\lambda t} \\
&= 1 - e^{-\lambda t} - \lambda t e^{-\lambda t}
\end{aligned}
$$

Now differentiate with respect to t to get the density:

$$f_{T_0}(t) = \lambda^2 t \exp(-\lambda t), t > 0$$

Notes

A couple things:

- this is a Gamma$(2, 1/\lambda)$ distribution.
- Recall that there are a few tricks that make Gamma distributions nice
- The distribution of sums of independent random variables is made easier by moment generating functions (we'll talk about these later)
- make sure you get a feel for this (a lot of the same tricks will reappear later)

Example

Extending example 6.3:

Let $X_1, X_2, \ldots, X_N \overset{iid}{\sim}$ Exponential(λ). What is the distribution of $T_0 = \sum_{i=1}^{n} X_i$? First we'll find it's cumulative distribution function (cdf) $F_{T_0}(t)$; then we'll take the derivative to get the density $f_{T_0}(t)$.

Example

$$
\begin{aligned}
F_{T_0}(t) &= P\left(\sum_{i=1}^{n} X_i \leq t\right) \\
&= \iint_{\{(x_1,x_2):x_1+x_2\leq t\}} f(x_1,\ldots,x_n)dx_{1:n} \\
&= \int_0^t \int_0^{t-\sum_{i=1}^{1} x_i} \cdots \int_0^{t-\sum_{i=1}^{n-1} x_i} f(s_1,\ldots,s_n)ds_n ds_{n-1}\cdots ds_1
\end{aligned}
$$

Example

Alternatively, you could use the transformation theorem:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_1 \\ X_1 + X_2 \\ \vdots \\ \sum_{i=1}^n X_i \end{bmatrix}$$

It isn't that bad to get $g(y_1, y_2, \ldots, y_n)$, but it's difficult to integrate out $y_1, y_2, \ldots, y_{n-1}$

We can also run simulations to check our derivations.

Example: If $X_1, \ldots, X_{100} \overset{iid}{\sim} \text{Normal}(0, 1)$, what does the maximum $X_{(100)}$ follow?