Chapter 7

Continuous Distributions

In Chapter 5 you met your first example of a continuous distribution, the normal. Recall the general definition.

Densities

A random variable X is said to have a ***continuous distribution*** (on $\mathbb{R}$) with ***density function*** $f(\cdot)$ if

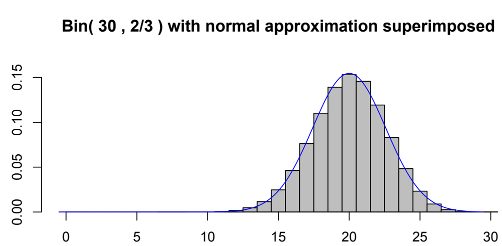
(i) f is a nonnegative function on the real line for which $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

(ii) for each subset A of the real line,

$$\mathbb{P}\{X \in A\} = \int_A f(x)\,dx = \int_{-\infty}^{\infty} \mathbb{I}\{x \in A\} f(x)\,dy$$

Assumption (ii) is actually equivalent to its special case:

$$\mathbb{P}\{a \le X \le b\} = \int_a^b f(x)\,dx \qquad \text{for all intervals } [a,b] \subseteq \mathbb{R}.$$



For the normal approximation to the Bin(n,p) the density was

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \qquad \text{for } -\infty < x < \infty$$

with $\mu = np$ and $\sigma^2 = npq$. That is, f is the $N(\mu, \sigma^2)$ density.

> **Remark.** As you will soon learn, the $N(\mu, \sigma^2)$ distribution has expected value μ and variance σ^2.

Notice that a change of variable $y = (x - \mu)/\sigma$ gives

$$\int_{-\infty}^{\infty} f(x)\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2}\,dy,$$

which (see Chapter 5) equals 1.

The simplest example of a continuous distribution is the ***Uniform***$[0,1]$, the distribution of a random variable U that takes values in the interval $[0,1]$, with

$$\mathbb{P}\{a \le U \le b\} = b - a \qquad \text{for all } 0 \le a \le b \le 1.$$

Equivalently,

$$\mathbb{P}\{a \le U \le b\} = \int_a^b f(x)\,dx \qquad \text{for all real } a, b,$$

where

$$f(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

I will use the Uniform to illustrate several general facts about continuous distributions.

> **Remark.** Of course, to actually simulate a Uniform$[0,1]$ distribution on a computer one would work with a discrete approximation. For example, if numbers were specified to only 7 decimal places, one would be approximating Uniform[0,1] by a discrete distribution placing probabilities of about 10^{-7} on a fine grid of about 10^7 equi-spaced points in the interval. You might think of the Uniform$[0,1]$ as a convenient idealization of the discrete approximation.

Be careful not to confuse the density $f(x)$ with the probabilities $p(y) = \mathbb{P}\{Y = y\}$ used to specify ***discrete distributions***, that is, distributions for random variables that can take on only a finite or countably infinite set of different values. The Bin(n,p) and the geometric(p) are both discrete distributions. Continuous distributions smear the probability out over a continuous range of values. In particular, if X has a continuous distribution with density f then

$$\mathbb{P}\{X = t\} = \int_t^t f(x)\,dx = 0 \qquad \text{for each fixed } t.$$

The value $f(x)$ does not represent a probability. Instead, the values taken by the density function could be thought of as constants of proportionality. At least at points where the density function f is continuous and when δ is small,

$$\mathbb{P}\{t \le X \le t+\delta\} = \int_t^{t+\delta} f(x)\,dy = f(t)\delta + \text{ terms of order } o(\delta).$$

Remark. Remember that $g(\delta) = o(\delta)$ means that $g(\delta)/\delta \to 0$ as $\delta \to 0$.

Equivalently,

$$\lim_{\delta\to 0} \frac{1}{\delta}\mathbb{P}\{t \le X \le t+\delta\} = f(t).$$

Some texts define the density as the derivative of the ***cumulative distribution function***

$$F(t) = \mathbb{P}\{-\infty < X \le t\} \qquad \text{for } -\infty < t < \infty.$$

That is,

$$f(t) = \lim_{\delta\to 0} \frac{1}{\delta}\big(F(t+\delta) - F(t)\big)$$

This approach works because

$$\begin{aligned} &\mathbb{P}\{t \le X \le t+\delta\} \\ &\qquad = \mathbb{P}\{X \le t+\delta\} - \mathbb{P}\{X < t\} \\ &\qquad = F(t+\delta) - F(t) \qquad \text{because } \mathbb{P}\{X = t\} = 0. \end{aligned}$$

Remark. Evil probability books often refer to random variables X that have continuous distributions as "continuous random variables", which is misleading. If you are thinking of a random variable as a function defined on a sample space, the so-called continuous random variables need not be continuous as functions.

> Evil probability books often also explain that distributions are called continuous if their distribution functions are continuous. A better name would be non-atomic: if X has distribution function F and if F has a jump of size p at x then $\mathbb{P}\{X = x\} = p$. Continuity of F (no jumps) implies no atoms, that is, $\mathbb{P}\{X = x\} = 0$ for all x. It is sad fact of real analysis life that continuity of F does not imply that the corresponding distribution is given by a density. Fortunately, you won't be meeting such strange beasts in this course.

When we are trying to determine a density function, the trick is to work with very small intervals, so that higher order terms in the lengths of the intervals can be ignored. (More formally, the errors in approximation tend to zero as the intervals shrink.)

Example <7.1> The distribution of $\tan(X)$ if $X \sim \text{Uniform}(-\pi/2, \pi/2)$

I recommend that you remember the method used in the previous Example, rather than trying to memorize the result for various special cases. In each particular application, rederive. That way, you will be less likely to miss multiple contributions to a density.

Example <7.2> Smooth functions of a random variable with a continuous distribution

---

Calculations with continuous distributions typically involve integrals or derivatives where discrete distribution involve sums or probabilities attached to individual points. The formulae developed in previous chapters for expectations and variances of random variables have analogs for continuous distributions.

Example <7.3> Expectations of functions of a random variable with a continuous distribution

---

You should be very careful not to confuse the formulae for expectations in the discrete and continuous cases. Think again if you find yourself integrating probabilities or summing expressions involving probability densities.

Example <7.4> Expected value and variance for the $N(\mu, \sigma^2)$.

Calculations for continuous distributions are often simpler than analogous calculations for discrete distributions because we are able to ignore some pesky cases.

Example <7.5> Zero probability for ties with continuous distributions.

Calculations are also greatly simplified by the fact that we can ignore contributions from higher order terms when working with continuous distributions and small intervals.

Example <7.6> The distribution of the order statistics from the uniform distribution.

The distribution from the previous Example is a member of a family whose name is derived from the ***beta function***, defined by

$$B(\alpha,\beta) := \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt \qquad \text{for } \alpha > 0, \beta > 0.$$

The equality

$$\int_0^1 t^{k-1}(1-t)^{n-k}dt = \frac{(k-1)!(n-k)!}{n!},$$

noted at the end of the Example, gives the value for $B(k, n-k+1)$.

In general, if we divide $t^{\alpha-1}(1-t)^{\beta-1}$ by $B(\alpha,\beta)$ we get a candidate for a density function: non-negative and integrating to 1.

Definition. For $\alpha > 0$ and $\beta > 0$ the Beta(α,β) distribution is defined by the density function

$$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \qquad \text{for } 0 < x < 1.$$

The density is zero outside $(0,1)$.

As you just saw in Example <7.6>, the kth order statistic from a sample of n independently generated random variables with Uniform$[0,1]$ distributions has a Beta$(k, n-k+1)$ distribution.

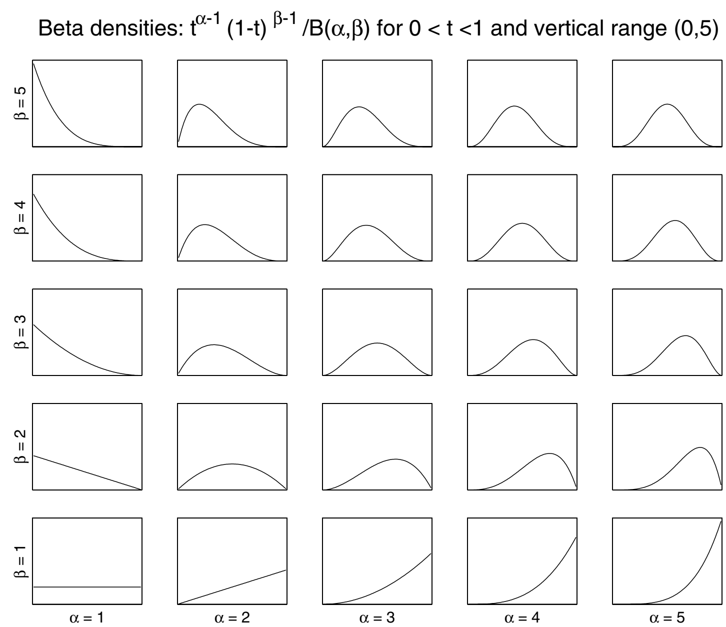
The function *beta()* in R calculates the value of the beta function:

```
> beta(5.5,2.7)
[1] 0.01069162
> ?beta   # get help for the beta() function
```

Also, there is a set of R functions that gives useful results for the beta density. For example, the pictures on the next page could be drawn by a series of R commands like:

```
> jj=(1:1000)/1000
> plot(jj,dbeta(jj,2,3),type="l")
```

The functions *dbeta()* calculates the values of the beta density at a fine grid of points. The *plot()* function is called with the option of joining the points by a smooth curve.



There is an interesting exact relationship between the tails of the beta and Binomial distributions.

Example <7.7> Binomial tail probabilities from beta distributions.

7.1 Things to remember

- The density function $f(\cdot)$ gives the constants of proportionality, and not probabilities: $f(x)$ is not the same as $\mathbb{P}\{X = x\}$, which is zero for every x if X has a continuous distribution.

- A density function, f, must be non-negative and it must integrate to one over the whole line, $1 = \int_{-\infty}^{\infty} f(t)\,dt$.
- Expected value of a function of a random variable with a continuous distribution: if the distribution of X has density f then

$$\mathbb{E}H(X) = \int_{-\infty}^{\infty} H(x)f(x)\,dx$$

- Be very careful not to confuse the formulae for expectations in the discrete and continuous cases. Think again if you find yourself integrating probabilities or summing expressions involving probability densities.

7.2 Examples for Chapter 7

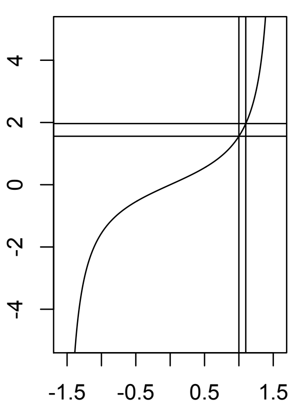
<7.1> **Example.** *The distribution of* $\tan(X)$ *if* $X \sim \text{Uniform}(-\pi/2, \pi/2)$

The distribution of X is continuous with density

$$f(x) = \mathbf{1}\{-\pi/2 < x < \pi/2\} = \begin{cases} 1/\pi & \text{for } -\pi/2 < x < \pi/2 \\ 0 & \text{elsewhere} \end{cases}$$

Let a new random variable be defined by $Y = \tan(X)$. It takes values over the whole real line. For a fixed real y, and a positive δ, we have

$$(*) \qquad y \le Y \le y + \delta \qquad \text{if and only if} \qquad x \le X \le x + \epsilon,$$



where x and ϵ are related to y and δ by the equalities

$$y = \tan(x) \qquad \text{AND} \qquad y + \delta = \tan(x + \epsilon).$$

By Calculus, for small δ,

$$\delta = (y + \delta) - y\epsilon \times \frac{\tan(x+\epsilon) - \tan(x)}{\epsilon} \approx \frac{\epsilon}{\cos^2 x}.$$

Compare with the definition of the derivative:

$$\lim_{\epsilon \to 0} \frac{\tan(x+\epsilon) - \tan(x)}{\epsilon} = \frac{d\tan(x)}{dx} = \frac{1}{\cos^2 x}.$$

Thus

$$\begin{aligned}\mathbb{P}\{y \le Y \le y+\delta\} &= \mathbb{P}\{x \le X \le x+\epsilon\} \\ &\approx \epsilon f(x) \\ &\approx \frac{\delta \cos^2 x}{\pi}.\end{aligned}$$

We need to express $\cos^2 x$ as a function of y. Note that

$$1+y^2 = 1+\frac{\sin^2 x}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}.$$

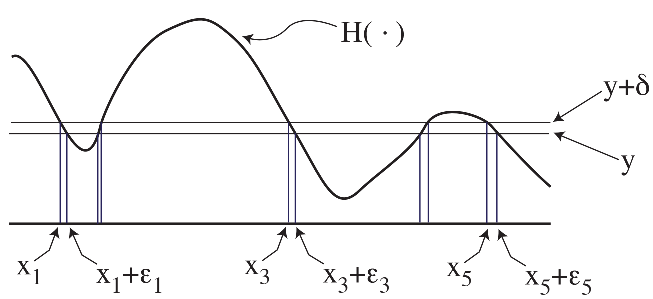
Thus Y has a continuous distribution with density

$$g(y) = \frac{1}{\pi(1+y^2)} \qquad \text{for } -\infty < y < \infty.$$

The continuous distribution with this density is called the ***Cauchy***. □

<7.2> **Example.** For functions that are not one-to-one, the analog of the method from Example <7.1> can require a little more work. In general, we can consider a random variable Y defined as $H(X)$, a function of another random variable. If X has a continous distribution with density f, and if H is a smooth function with derivative H', then we can calculate a density for Y by an extension of the method for the *tan* function.

A small interval $[y, y+\delta]$ in the range of values taken by Y can correspond to a more complicated range of values for X. For instance, it might consist of a union of several intervals $[x_1, x_1+\epsilon_1]$, $[x_2, x_2+\epsilon_2]$, ... The number of pieces in the X range might be different for different values of y.



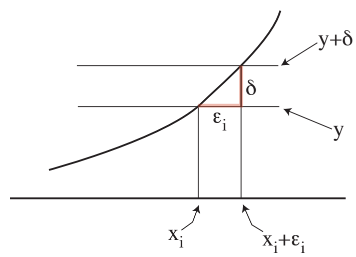
From the representation of $\{y \le Y \le y+\delta\}$ as a disjoint union of events

$$\{x_1 \le X \le x_1+\epsilon_1\} \cup \{x_2 \le X \le x_2+\epsilon_2\} \cup \ldots,$$

we get, via the defining property of the density f for X,

$$\begin{aligned}\mathbb{P}\{y \le Y \le y + \} &= \mathbb{P}\{x_1 \le X \le x_1 + \epsilon_1\} + \mathbb{P}\{x_2 \le X \le x_2 + \epsilon_2\} + \dots \\ &\approx \epsilon_1 f(x_1) + \epsilon_2 f(x_2) + \dots.\end{aligned}$$

For each small interval, the ratio of δ/ϵ_i is close to the derivative of the function H at the corresponding x_i. That is, $\epsilon_i \approx \delta/H'(x_i)$.



Adding the contributions from each such interval, we then have an approximation that tells us the density for Y,

$$\mathbb{P}\{y \le Y \le y + \delta\} \approx \delta \left(\frac{f(x_1)}{H'(x_1)} + \frac{f(x_2)}{H'(x_2)} + \dots \right)$$

That is, the density for Y at the particular point y in its range equals

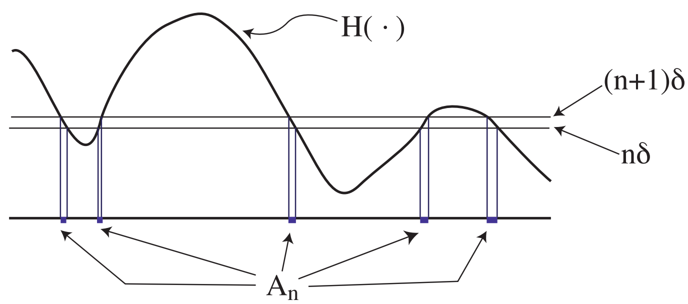
$$\frac{f(x_1)}{H'(x_1)} + \frac{f(x_2)}{H'(x_2)} + \dots$$

Of course we should reexpress each x_i as a function of y, to get the density in a more tractable form. □

<7.3> **Example.** *Expectations of functions of a random variable with a continuous distribution*

Suppose X has a continuous distribution with density function f. Let $Y = H(X)$ be a new random variable, defined as a function of X. We can calculate $\mathbb{E}Y$ by an approximation argument similar to the one used in Example <7.2>. It will turn out that

$$\mathbb{E}H(X) = \int_{-\infty}^{+\infty} H(x) f(x)\, dx.$$



Cut the range of values that might be taken by Y into disjoint intervals of the form $n\delta \le y < (n+1)\delta$, for an arbitrarily small, positive δ. Write A_n for the corresponding set of x values. That is, for each x in $\mathbb{R}$,

$$n\delta \le H(x) < (n+1)\delta \qquad \text{if and only if} \qquad x \in A_n.$$

We now have simple upper and lower bounds for H:

$$H_\delta(x) \le H(x) \le \delta + H_\delta(x) \qquad \text{for every real } x$$
$$\text{where } H_\delta(x) = \sum\nolimits_n n\delta \mathbf{1}\{x \in A_n\}.$$

(You should check the inequalities when $x \in A_n$, for each possible integer n.) Consequently

$$\mathbb{E}H_\delta(X) \le \mathbb{E}H(X) \le \delta + \mathbb{E}H_\delta(X)$$

and

$$\int_{-\infty}^{+\infty} H_\delta(x)f(x)\,dx \le \int_{-\infty}^{+\infty} H(x)f(x)\,dx \le \delta + \int_{-\infty}^{+\infty} H_\delta(x)f(x)\,dx.$$

More concisely,

$$(\star) \qquad |\mathbb{E}H_\delta(X) - \mathbb{E}H(X)| \le \delta$$

and

$$(\star\star) \qquad |\int_{-\infty}^{+\infty} H_\delta(x)f(x)\,dx - \int_{-\infty}^{+\infty} H(x)f(x)\,dx| \le \delta.$$

The random variable $H_\delta(X)$ has a discrete distribution whose expectation you know how to calculate:

$$\begin{aligned}
\mathbb{E}H_\delta(X) &= \mathbb{E}\sum\nolimits_n n\delta \mathbf{1}\{X \in A_n\} && \text{expectation of a countable sum} \\
&= \sum\nolimits_n n\delta \mathbb{P}\{X \in A_n\} && \text{because } \mathbb{E}\mathbf{1}\{X \in A_n\} = \mathbb{P}\{X \in A_n\} \\
&= \sum\nolimits_n n\delta \int_{-\infty}^{+\infty} \mathbf{1}\{x \in A_n\} f(x)\,dx && \text{definition of } f \\
&= \int_{-\infty}^{+\infty} H_\delta(x) f(x)\,dx && \text{take sum inside integral.}
\end{aligned}$$

From the inequalities ($\star$) and ($\star\star$), the last equality deduce that

$$|\mathbb{E}H(X) = \int_{-\infty}^{+\infty} H(x) f(x)\,dx| \le 2\delta$$

for arbitrarily small $\delta > 0$. The asserted representation for $\mathbb{E}H(X)$ follows. $\square$

Remark. Compare with the formula for a random variable X^* taking only a discrete set of values $x_1, x_2, \dots$,

$$\mathbb{E}H(X^*) = \sum\nolimits_i H(x_i)\mathbb{P}\{X^* = x_i\}$$

In the passage from discrete to continuous distributions, discrete probabilities get replaced by densities and sums get replaced by integrals.

<7.4> **Example.** *Expected value and variance* $N(\mu, \sigma^2)$.

If $X \sim N(\mu, \sigma^2)$ its density function is

$$\begin{aligned}
f(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) && \text{for } -\infty < x < \infty \\
&= \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) && \text{where } \phi(y) := (2\pi)^{-1/2}\exp(-y^2/2).
\end{aligned}$$

Calculate, using a change of variable $y = (x-\mu)/\sigma$.

$$\begin{aligned}
\mathbb{E}X &= \int_{-\infty}^{+\infty} x f(x)\,dx \\
&= \int_{-\infty}^{+\infty} (\mu + \sigma y)\phi(y)\,dy \\
&= \mu \int_{-\infty}^{+\infty} \phi(y)\,dy + \sigma \int_{-\infty}^{+\infty} y\phi(y)\,dy \\
&= \mu.
\end{aligned}$$

The second integral vanishes because $y\phi(y) = -(-y)\phi(-y)$.

Similarly

$$\begin{aligned} \text{var}(X) &= \mathbb{E}(X-\mu)^2 \\ &= \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx \\ &= \sigma^2 \int_{-\infty}^{+\infty} y^2 \phi(y)\,dy \\ &= \sigma^2 \end{aligned}$$

using integration by parts and $\frac{d}{dy}\phi(y) = -y\phi(y)$.

□

<7.5> **Example.** Suppose X and Y are independent random variables, each with a Uniform$[0,1]$ distribution. Show that $\mathbb{P}\{X = Y\} = 0$.

The event $\{X = Y = 1\}$ is a subset of $\{X = 1\}$, which has zero probability. The other possibilities are almost as easy to dispose of: for each positive integer n,

$$\{X = Y < 1\} \subset \cup_{i=0}^{n-1} \{i/n \le X < (i+1)/n \text{ and } i/n \le Y < (i+1)/n\}$$

a disjoint union of events each with probability $1/n^2$, by independence. Thus

$$\mathbb{P}\{X = Y < 1\} \le n(1/n^2) = 1/n \qquad \text{for every } n.$$

It follows that $\mathbb{P}\{X = Y\} = 0$.

A similar calculation shows that $\mathbb{P}\{X = Y\} = 0$ for independent random variables with any pair of continuous distributions. □

<7.6> **Example.** *The distribution of the order statistics from the uniform distribution.*

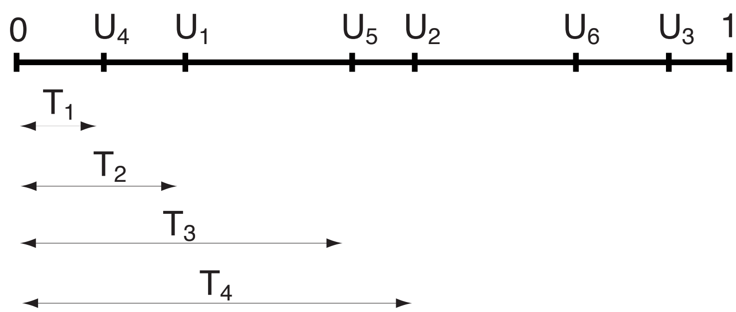
Suppose $U_1, U_2, \dots, U_n$ are independent random variables, each with distribution Uniform$(0,1)$. That is,

$$\mathbb{P}\{a \le U_i \le b\} = \int_a^b h(x)\,dx \qquad \text{for all real } a \le b,$$

where

$$h(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The U_i's define n points in the unit interval. If we measure the distance of each point from 0 we obtain random variables $0 \le T_1 < T_2 < \cdots < T_n \le 1$, the values $U_1, \dots, U_n$ rearranged into increasing order. (Example <7.5> lets me ignore ties.) For $n = 6$, the picture (with T_5 and T_6 not shown) looks like:



If we repeated the process by generating a new sample of U_i's, we would probably not have U_4 as the smallest, U_1 as the second smallest, and so on. That is, T_1 might correspond to a different U_i.

The random variable T_k, the kth smallest of the ordered values, is usually called the kth ***order statistic***. It takes a continuous range of values. It has a continuous distribution. What is its density function?

For a very short interval $[t, t+\delta]$, with $0 < t < t+\delta < 1$ and δ very small, we need to show that $\mathbb{P}\{t \le T_k \le t+\delta\}$ is roughly proportional to δ, then determine $f(t)$, the constant of proportionality.

Write N for the number of U_i points that land in $[t, t+\delta]$. To get $t \le T_k \le t+\delta$ we must have $N \ge 1$. If $N = 1$ then we must have exactly $k-1$ points in $[0, t)$ to get $t \le T_k \le t+\delta$. If $N \ge 2$ then it becomes more complicated to describe all the ways that we would get $t \le T_k \le t+\delta$. Luckily for us, the contributions from all those complicated expressions will turn out to be small enough to ignore if δ is small. Calculate.

$$\begin{aligned}\mathbb{P}\{t \le T_k \le t+\delta\} &= \mathbb{P}\{N = 1 \text{ and exactly } k-1 \text{ points in } [0, t)\}\\ &\qquad + \mathbb{P}\{N \ge 2 \text{ and } t \le T_k \le t+\delta\}.\end{aligned}$$

Let me first dispose of the second contribution, where $N \ge 2$. The event

$$F_2 = \{N \ge 2\} \cap \{t \le T_k \le t+\delta\}$$

is a subset of the union

$$\cup_{1<i<j\le n}\{U_i,\, U_j \text{ both in } [t, t+\delta]\,\}$$

Put another way,

$$\mathbb{I}_{F_2} \le \sum\nolimits_{1\le i<j\le n} \mathbb{I}\{U_i,\, U_j \text{ both in } [t, t+\delta]\,\}.$$

Take expectations of both sides to deduce that

$$\mathbb{P}F_2 \le \sum\nolimits_{1\le i<j\le n} \mathbb{P}\{U_i,\, U_j \text{ both in } [t, t+\delta]\}.$$

By symmetry, all $\binom{n}{2}$ terms in the sum are equal to

$$\begin{aligned}
&\mathbb{P}\{U_1, U_2 \text{ both in } [t, t+\delta]\} \\
&\qquad = \mathbb{P}\{t \le U_1 \le t+\delta\}\mathbb{P}\{t \le U_2 \le t+\delta\} \qquad \text{by independence} \\
&\qquad = \delta^2.
\end{aligned}$$

Thus $\mathbb{P}F_2 \le \binom{n}{2}\delta^2$, which tends to zero much faster than δ as $\delta \to 0$. (The value of n stays fixed throughout the calculation.)

Next consider the contribution from the event

$$F_1 = \{N = 1\} \cap \{\text{exactly } k-1 \text{ points in } [0, t)\}.$$

Break F_1 into disjoint events like

$$\{U_1, \dots, U_{k-1} \text{ in } [0, t),\, U_k \text{ in } [t, t+\delta],\, U_{k+1}, \dots, U_n \text{ in } (t+\delta, 1]\}.$$

Again by virtue of the independence between the $\{U_i\}$, this event has probability

$$\begin{aligned}
&\mathbb{P}\{U_1 < t\}\mathbb{P}\{U_2 < t\} \dots \mathbb{P}\{U_{k-1} < t\} \\
&\qquad\qquad \times \mathbb{P}\{U_k \text{ in } [t, t+\delta]\} \\
&\qquad\qquad \times \mathbb{P}\{U_{k+1} > t+\delta\} \dots \mathbb{P}\{U_n > t+\delta\},
\end{aligned}$$

Invoke the defining property of the uniform distribution to factorize the probability as

$$t^{k-1}\delta(1-t-\delta)^{n-k} = t^{k-1}(1-t)^{n-k}\delta + \text{ terms of order } \delta^2 \text{ or smaller.}$$

How many such pieces are there? There are $\binom{n}{k-1}$ ways to choose the $k-1$ of the U_i's to land in $[0, t)$, and for each of these ways there are $n-k+1$ ways to choose the single observation to land in $[t, t+\delta]$. The remaining observations must go in $(t+\delta, 1]$. We must add up

$$\binom{n}{k-1} \times (n-k+1) = \frac{n!}{(k-1)!(n-k)!}$$

contributions with the same probability to calculate $\mathbb{P}F_1$.

Consolidating all the small contributions from $\mathbb{P}F_1$ and $\mathbb{P}F_2$ we then get

$$\begin{aligned}&\mathbb{P}\{t \le T_k \le t+\delta\}\\ &\quad = \frac{n!}{(k-1)!(n-k)!}t^{k-1}(1-t)^{n-k}\delta + \text{ terms of order } \delta^2 \text{ or smaller.}\end{aligned}$$

That is, the distribution of T_k is continuous with density function

$$f(t) = \frac{n!}{(k-1)!(n-k)!}t^{k-1}(1-t)^{n-k} \qquad \text{for } 0 < t < 1.$$

Outside $(0,1)$ the density is zero. □

Remark. It makes no difference how we define $f(t)$ at $t = 0$ and $t = 1$, because it can have no effect on integrals $\int_a^b f(t)\,dt$.

From the fact that the density must integrate to 1, we get

$$1 = \int_{-\infty}^{0} 0dt + \frac{n!}{(k-1)!(n-k)!}\int_0^1 t^{k-1}(1-t)^{n-k}dt + \int_1^{\infty} 0dt$$

That is,

$$\int_0^1 t^{k-1}(1-t)^{n-k}dt = \frac{(k-1)!(n-k)!}{n!},$$

a fact that you might try to prove by direct calculation.

<7.7> **Example.** *Binomial tail probabilities from beta distributions.*

In principle it is easy to calculate probabilities such as $\mathbb{P}\{\text{Bin}(30,p) \ge 17\}$ for various values of p: one has only to sum the series

$$\binom{30}{17}p^{17}(1-p)^{13} + \binom{30}{18}p^{18}(1-p)^{12} + \cdots + (1-p)^{30}.$$

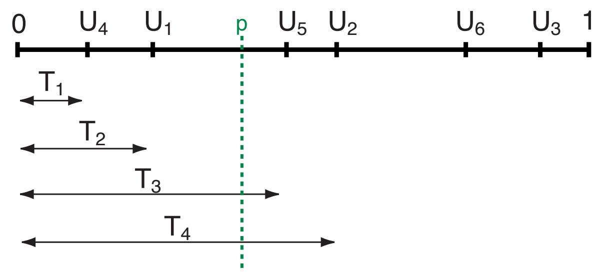
With a computer (using R, for example) such a task would not be as arduous as it used to be back in the days of hand calculation. We could also use a normal approximation (as in the example for the median in Chapter 6). However, there is another method based on the facts about the order statistics, which gives an exact integral expression for the Binomial tail probability.

The relationship becomes clear from a special method for simulating coin tosses. For a fixed n (such as $n = 30$), generate independently n random variables $U_1, \dots, U_n$, each distributed uniformly on $[0, 1]$. Fix a p in $[0, 1]$. Then the independent events

$$\{U_1 \le p\}, \{U_2 \le p\}, \dots, \{U_n \le p\}$$

are like n independent flips of a coin that lands heads with probability p. The number, X_n, of such events that occur has a $\text{Bin}(n, p)$ distribution.

As in Example <7.6>, write T_k for the kth smallest value when the U_i's are sorted into increasing order.



The random variables X_n and T_k are related by an equivalence,

$$X_n \ge k \text{ if and only if } T_k \le p.$$

That is, there are k or more of the U_i's in $[0, p]$ if and only if the kth smallest of all the U_i's is in $[0, p]$. Thus

$$\mathbb{P}\{X_n \ge k\} = \mathbb{P}\{T_k \le p\} = \frac{n!}{(k-1)!(n-k)!} \int_0^p t^{k-1}(1-t)^{n-k}\,dt.$$

The density for the distribution of T_k comes from Example <7.6>. □