14.30 Introduction to Statistical Methods in Economics
Spring 2009

14.30 Introduction to Statistical Methods in Economics Lecture Notes 11

Konrad Menzel

March 17, 2009

1 Order Statistics

Let $X_1, \ldots, X_n$ be independent random variables with identical p.d.f.s $f_{X_1}(x) = \ldots = f_{X_n}(x)$ - we'll generally call such a sequence "independent and identically distributed", which is typically abbreviated as *iid*. We are interested in the function

$$Y_n = \max\{X_1, \ldots, X_n\}$$

i.e. Y_n is the largest value in the sample.

We can derive the c.d.f. of Y_n using independence

$$\begin{aligned}
F_{Y_n}(y) &= P(Y_n \leq y) = P(X_1 \leq y, X_2 \leq y, \ldots, X_n \leq y) \\
&= P(X_1 \leq y)P(X_2 \leq y)\ldots P(X_n \leq y) \\
&= F_{X_1}(y)F_{X_2}(y)\ldots F_{X_n}(y) \\
&= [F_X(y)]^n
\end{aligned}$$

Using the chain rule, we can obtain the p.d.f. of the maximum,

$$f_{Y_n}(y) = \frac{d}{dy}F_{Y_n}(y) = n\left[F_X(y)\right]^{n-1}f_X(y)$$

Example 1 *An old painting is sold at an auction. n identical bidders submit their bids $B_1, \ldots, B_n$ independently, and the marginal c.d.f. of the bids is $F_B(b)$. The potential buyer who submitted the highest bid gets to buy the painting and has to pay his bid (this type of auction is called Dutch, or first price auction). Then the revenue of the seller of the painting has p.d.f. is*

$$f_Y(y) = f_{\max\{B_1, \ldots, B_n\}}(y) = n\left[F_B(y)\right]^{n-1}f_B(y)$$

Now we can generalize this to other ranks in the sample, e.g.

$$Y_{n-1} = \text{"The 2nd highest value in } X_1, \ldots, X_n \text{"}$$

This random variable is called the $(n-1)th$ *order statistic* of $X_1, \ldots, X_n$, and we can state its p.d.f.

Proposition 1 *Let $X_1, \ldots, X_n$ be an iid sequence of random variables with p.d.f. $f_X(f)$ and c.d.f. $F_X(x)$. Then the kth order statistic Y_k has p.d.f.*

$$f_{Y_k}(y) = k \left(\begin{array}{c} n \\ k \end{array}\right) [F_X(y)]^{k-1}[1 - F_X(y)]^{n-k}f_X(y)$$

PROOF: We can split the experiment into two parts, (a) one of the X_i's has to take the value y according to the density $f_X(y)$, and (b) given the value y, the other draws in the sequence have to be grouped around y in a way that makes y the kth smallest value in the sample.

Part (b) is a binomial experiment in which the n trials correspond to the n draws of $X_1, \ldots, X_n$, and we define a "success" in the ith round as the event $(X_i \leq y)$. Since draws are independent and correspond to the same p.d.f. the parameter p in the binomial distribution is equal to $F_X(y)$. y being smaller or equal to the kth smallest value corresponds to *at least* k "successes" in the binomial experiment, and therefore the corresponding c.d.f. is

$$F_{Y_k}(y) = \sum_{l=k}^{n} \binom{n}{l} [F_X(y)]^l [1 - F_X(y)]^{n-l}$$

We can now obtain the p.d.f. by differentiating the c.d.f. with respect to y, using the product and the chain rule

$$
\begin{aligned}
f_{Y_k}(y) &= \frac{d}{dy} F_{Y_k}(y) \\
&= \sum_{l=k}^{n} \binom{n}{l} l[F_X(y)]^{l-1}[1 - F_X(y)]^{n-l} f_X(y) - \sum_{l=k}^{n} \binom{n}{l} (n-l)[F_X(y)]^l [1 - F_X(y)]^{n-l-1} f_X(y) = T_1 - T_2
\end{aligned}
$$

This expression looks complicated, but it turns out to be essentially a telescopic sum, so that most summands will drop out. Noting that for the second term, the summand corresponding to $l = n$ is zero, we can rewrite it as

$$T_2 = \sum_{l=k}^{n-1} \binom{n}{l} (n-l)[F_X(y)]^l [1-F_X(y)]^{n-l-1} f_X(y) = \sum_{l=k+1}^{n} \binom{n}{l-1} (n-l+1)[F_X(y)]^{l-1}[1-F_X(y)]^{n-l} f_X(y)$$

replacing the running index l with $l-1$. Since for the first term,

$$l \cdot \binom{n}{l} = \frac{n!l}{l!(n-l)!} = \frac{n!}{(l-1)!(n-l+1)!}(n-l+1) = \binom{n}{l-1}(n-l+1)$$

The first term becomes

$$T_1 = \sum_{l=k}^{n} \binom{n}{l-1} (n-l+1)[F_X(y)]^{l-1}[1 - F_X(y)]^{n-l} f_X(y)$$

so that the density equals the $l = k$ term of the sum defining T_1 since it is the only one which doesn't get canceled out when we subtract T_2. Therefore

$$
\begin{aligned}
f_{Y_k}(y) &= T_1 - T_2 = \sum_{l=k}^{k} \binom{n}{l-1} (n-l+1)[F_X(y)]^{l-1}[1 - F_X(y)]^{n-l} f_X(y) \\
&= \frac{n!}{k!(n-k+1)!}(n-k+1)[F_X(y)]^{k-1}[1 - F_X(y)]^{n-k} f_X(y) \\
&= k \binom{n}{k} [F_X(y)]^{k-1}[1 - F_X(y)]^{n-k} f_X(y)
\end{aligned}
$$

which is the result we were going to prove $\square$

Example 2 *We could now think of an auction which is different from the first one in that the buyer submitting the highest bid still gets to buy the painting, but only has to pay the amount offered by the second-highest bidder (this auction format is more common than the first, and it's known as an English, or second-price auction). If the submitted bids are random variables $C_1, \ldots, C_n$, the revenue Y of the seller now has p.d.f.*

$$f_Y(y) = 2 \begin{pmatrix} n \\ 2 \end{pmatrix} F_C(y)[1 - F_C(y)]^{n-2} f_C(y)$$

Notice that I used different letters for the bids, since it is known from economic theory that the same bidders should submit different bids under the two different auction formats.

2 Digression: The Distribution of the "First Digit of Anything" (skipped in lecture)

Here's a somewhat cute but non-standard problem, for which we are *not* going to use the methods we saw earlier. So this is definitely something you can skip for problem sets or exam preparation, but I'd still like to go over it.

What is the distribution of the first digit of a number we basically don't know anything about? To be more precise, let X be a measurement of any kind, for which we don't know what it represents, nor in what units $\frac{1}{Y}$ it is measured - e.g. we could leaf through a newspaper and collect any numbers representing a measurement of some kind (incomes, stock indices, population etc.). What is the p.d.f. of the first digit of a number of that kind if we don't know anything else about where it comes from? I.e. how do we derive the p.d.f. of the first decimal of the random variable $Z = X \cdot Y$ if all we know that X and Y are positive, but can be anything?

A first guess might be that the distribution of the first digit is (discrete) uniform, since this distribution intuitively does not seem to contain much "information" about the numbers nor their units. However, if we take a uniform distribution and change the units (e.g. double or quadruple all numbers in the assumed underlying distribution), the distribution of first digits doesn't remain uniform. E.g. if the true numbers are $X \sim U[1, 10]$, $4X \sim U[0, 40]$, so that for the first digit Y of $4X$ has p.d.f.

$$f_Y(y) = \begin{cases} \frac{1}{40} + \frac{1}{4} & \text{if } y \in \{1, 2, 3\} \\ \frac{1}{40} & \text{if } y \in \{4, 5, 6, 7, 8, 9\} \\ 0 & \text{otherwise} \end{cases}$$

Or, in pictures, So a minimal requirement for the distribution of the first digits should be that it doesn't change if we change the units of measurement.

What we are looking for is in fact a random variable X for which the distribution of X doesn't change for changes of the scale, i.e. aX for $a > 0$. This is true if we assume that $Z = \log(X) \sim U[\log(1), \log(10)]$, since for a scale shift,

$$\begin{aligned} P(z \leq aX \leq z + 1) &= P\left(\frac{z}{a} \leq X \leq \frac{z+1}{a}\right) = \frac{\log\left(\frac{z}{a}\right) - \log\left(\frac{z+1}{a}\right)}{\log(10a) - \log(a)} \\ &= \frac{\log(z+1) - \log(z)}{\log(10) - \log(1)} = P(z \leq X \leq z + 1) \end{aligned}$$

Then the first digit Y of Z has p.d.f.

$$f_Y(y) = \begin{cases} \frac{\log(z+1) - \log(z)}{\log(10)} & \text{if } y \in \{1, 2, 3, \ldots, 9\} \\ 0 & \text{otherwise} \end{cases}$$

3
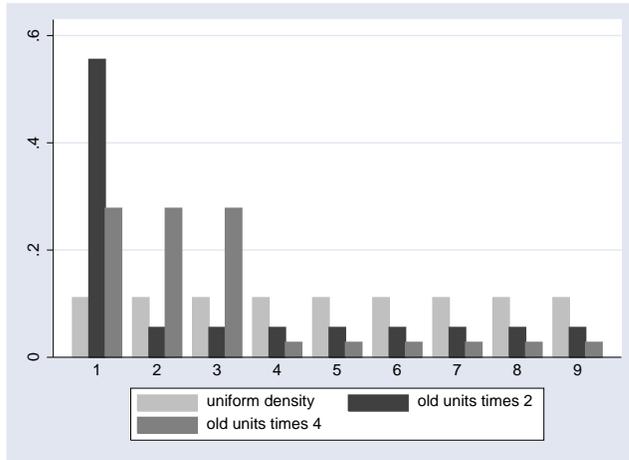
Figure 1: Effect of a Change of Measurement Units on the Uniform

This invariance idea may seem like a very artificial way of obtaining a distribution, since there is no obvious connection to the measurements X nor the measurement units. However, the resulting p.d.f. seems to give a good approximation to "real-world data" which falls into the category - as e.g. numbers representing measurements of some kind which appear in the New York Times. The following graph shows the "theoretical" density together with a histogram of the first digit of GDP measured in local currency units (i.e. Yen for Japan, C\$ for Canada, etc.) for the 77 countries included in "World in Figures 2007" pocket book from The Economist. Summarizing, this example took a radically different approach
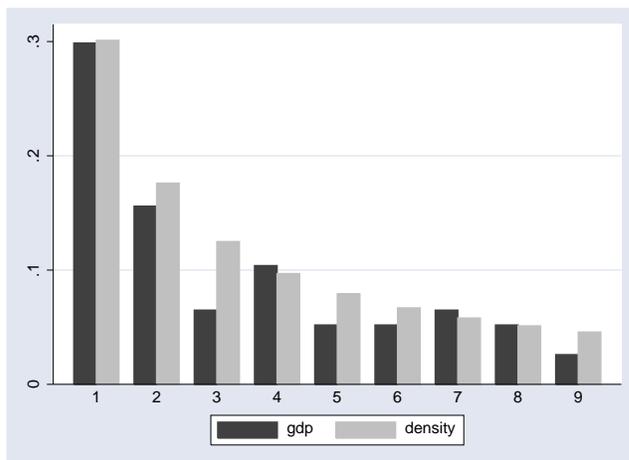
Figure 2: Distribution of First Digit of GDP in Local Currency Units and Theoretical Density (Numbers from The Economist, Pocket World in Figures 2007)

to determining the distribution of a given function of two random variables X and Y: here we started out not knowing the p.d.f.s of X and Y, but then imposed that whatever the resulting distribution was going to be, it had to be *invariant* with respect to a change of units - i.e. the realization of Y. The concept of "invariance" plays a major role in advanced statistics, but for the purpose of this class, we

won't go beyond this example.

3 Expected Values and Median

Given a random variable X with a p.d.f. $f_X(x)$, we'd like to summarize the most important characteristics of the entire distributions without having to give the entire density function. The expected value tells us basically where the distribution of X is centered.

3.1 Definitions

Definition 1 *If X is a discrete random variable, the expected value of X, denoted $\mathbb{E}[X]$ is given by*

$$\mathbb{E}[X] := \sum_x x f_X(x)$$

if this sum is finite. If X is continuous, the expected value is defined as

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

if the integral is finite.

Example 3 *What is the expectation of a binomial random variable, i.e. $X \sim B(n, p)$?*

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n! x}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} np \frac{(n-1)!}{(x-1)!((n-1)-(x-1))!} p^{x-1} (1-p)^{(n-1)-(x-1)} \\
&= np \sum_{x=0}^{n-1} \binom{n-1}{x-1} p^x (1-p)^{n-1-x} \\
&= np \cdot 1
\end{aligned}
$$

where in the second row, we can ignore the summand corresponding to $x = 0$ since it's zero, in the third row, we pull n out of the binomial coefficient, and in the following step, we switch the summation index from x to $x-1$. Finally, if we pull np out, the summands are the binomial probabilities for $\tilde{X} \sim B(n-1, p)$, and therefore they sum to one.

Note that a random variable which takes infinitely many different value may not have a finite expectation, in which case the expectation is not defined. You should also notice that even though the expectation gives a sense of the "location" of a distribution, it is in general not a "typical" value for the random variable: e.g. the expected value of a die roll is $\frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3\frac{1}{2}$, which is not a possible outcome.

An alternative measure of the position of a distribution is the *median*:

Definition 2 *The* median *$m(X)$ of a random variable X is a real number such that*

$$P(X < m) = \frac{1}{2}$$

The median and the expected value of a random variable X coincide if the distribution of X is symmetric around $m(X)$, i.e. $f_X(m(X) - x) = f_X(m(x) + x)$, but need not be the same in general.

Example 4 *Say X has p.d.f.*

$$f_X(x) = \begin{cases} \frac{1}{9}x^2 & \text{if } 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

The expected value is

$$\int_0^3 t \cdot \frac{1}{9}t^2 dt = \frac{1}{9}\int_0^3 t^3 dt = \left[\frac{1}{36}t^4\right]_0^3 = \frac{81}{36} = \frac{9}{4} = 2.25$$

In order to obtain the median, let's first calculate the c.d.f. of X

$$F_X(x) = \int_0^x \frac{1}{9}t^2 dt = \left[\frac{1}{27}t^3\right]_0^x = \frac{x^3}{27}$$

Therefore, solving $F_X(m) = \frac{1}{2}$ for m gives

$$m^3 = \frac{27}{2} \Leftrightarrow m = \frac{3}{\sqrt[3]{2}} \approx 2.38 > 2.25$$

Therefore, the median of this distribution is greater than the mean.

Note that the median may not be unique:

Example 5 *Let X be the result of rolling a fair die once. Then for any number $m \in (3, 4]$, $P(X < m) = P(X \leq 3) = \frac{1}{2}$. Therefore any value in that interval is a median.*

3.2   Properties of Expectations

Property 1 *If $X = c$, where c is a constant, then*

$$\mathbb{E}[X] = c$$

Property 2 *If $Y = aX + b$, then*

$$\mathbb{E}[Y] = a\mathbb{E}[X] + b$$

PROOF: Let's only look at the continuous case: if X is a continuous random variable with p.d.f. $f_X(x)$, then the expectation of Y is given by

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty}(ax + b)f_X(x)dx = a\int_{-\infty}^{\infty}xf_X(x)dx + b\int_{-\infty}^{\infty}f_X(x)dx = a\mathbb{E}[X] + b \cdot 1$$

so as we can see, the linearity of integrals translates directly to linearity of expectations□

Property 3 *For $Y = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n + b$,*

$$\mathbb{E}[Y] = a_1\mathbb{E}[X_1] + a_2\mathbb{E}[X_2] + \ldots + a_n\mathbb{E}[X_n] + b$$

This is the most general statement of the linearity of expectations, and we will use this property over and over in the remainder of this class.

Example 6 *Above, we calculated the expectation of $X \sim B(n, p)$, $\mathbb{E}[X] = np$ by summing over the possible outcomes of X. But from the last result we can see that there is an easier way of obtaining the same result: since X is the number of successes from a sequence of n trials, we can code the outcome of each trial as $Z_1, Z_2, \ldots, Z_n$, where $Z_i = 1$ if the ith trial was a success and $Z_i = 0$ otherwise.*

$$\mathbb{E}[Z_i] = 1 \cdot p + 0 \cdot (1 - p) = p$$

and therefore,

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} Z_i\right] = \sum_{i=1}^{n} \mathbb{E}[Z_i] = \sum_{i=1}^{n} p = np$$

Property 4 *IF X and Y are* independent*, then*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

If X and Y are not independent, this will generally not be true.

3.3 Expectations of Functions of Random Variables

Let $Y = r(X)$. Last week, we saw how we could derive the p.d.f. of Y if we knew $f_X(x)$. For expectations, this problem is much simpler since we are only looking at one single characteristic of the distribution. The expectation of $Y = r(X)$ is given by

$$\mathbb{E}[Y] = \mathbb{E}[r(X)] = \begin{cases} \sum_x r(x)f_X(x) & \text{if } X \text{ discrete} \\ \int_{-\infty}^{\infty} r(t)f_X(t)dt & \text{if } X \text{ continuous} \end{cases}$$

Example 7 *Suppose $Y = X^{1/2}$ where the p.d.f. of X is given by*

$$f_X(x) = \begin{cases} 2x & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}[Y] = \int_0^1 t^{1/2} f_X(t)dt = 2\int_0^1 t^{3/2}dt = 2\left[\frac{2}{5}x^{5/2}\right]_0^1 = \frac{4}{5}$$

The same principle works for functions of 2 or more random variables

Example 8 *Suppose we are interested in the function $Z = X^2 + Y^2$ of two random variables X and Y with joint p.d.f.*

$$f_{XY}(x, y) = \begin{cases} 1 & \text{if } 0 \leq x, y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\begin{aligned}
\mathbb{E}[Z] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x^2 + y^2) f_{XY}(x, y)dxdy \\
&= \int_0^1 \int_0^1 (x^2 + y^2)dxdy \\
&= \int_0^1 \left[\frac{1}{3}x^3 + y^2 x\right]_0^1 dy \\
&= \int_0^1 \left(\frac{1}{3} + y^2\right) dy \\
&= \left[\frac{1}{3}y + \frac{1}{3}y^3\right]_0^1 = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}
\end{aligned}$$

For linear functions of a random variable X of the type $Y = aX + b$, we saw above that $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$. This doesn't work for *nonlinear* functions of a random variable. A particulary important result on this is *Jensen's Inequality*:
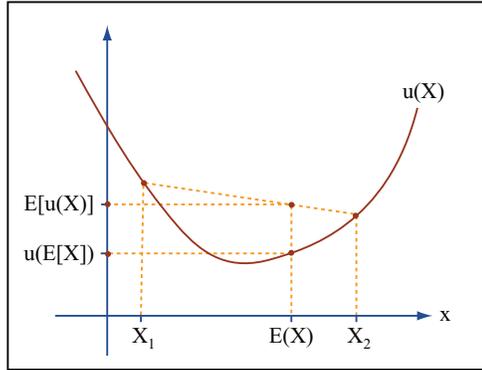
Image by MIT OpenCourseWare.

Figure 3: Example with a discrete distribution over $\{x_1, x_2\}$

Proposition 2 *(Jensen's Inequality) Let X be a random variable, and $u(x)$ be a convex function. Then*

$$\mathbb{E}[u(X)] \geq u\left(\mathbb{E}[X]\right)$$

The inequality is strict if $u(\cdot)$ is strictly *convex and X takes at least two different values with positive probability.*

PROOF: We can define a linear function

$$r(x) = u(\mathbb{E}[X]) + u'(\mathbb{E}[X])(x - \mathbb{E}[X])$$

which is tangent to $u(x)$ which passes through the point $(\mathbb{E}[X], u(\mathbb{E}[X]))$. Since $u(\cdot)$ is convex,

$$u(x) \geq r(x) \quad \text{for all } x$$

In particular,

$$\mathbb{E}[u(X)] = \int_{-\infty}^{\infty} u(t)f_X(t)dt \geq \int_{-\infty}^{\infty} r(t)f_X(t)dt = \mathbb{E}[r(X)]$$

Since $r(x)$ is linear by construction, we can use property 2 on expectations of linear functions with $a = u'(\mathbb{E}[X])$ and $b = u(\mathbb{E}[X]) - u'(\mathbb{E}[X])\mathbb{E}[X]$ to obtain

$$\mathbb{E}[r(X)] = \mathbb{E}[aX + b] = a\mathbb{E}[X] + b = u'(\mathbb{E}[X])\mathbb{E}[X] + u(\mathbb{E}[X]) - u'(\mathbb{E}[X])\mathbb{E}[X] = u(\mathbb{E}[X])$$

Putting this together with the inequality derived before,

$$\mathbb{E}[u(X)] \geq \mathbb{E}[r(X)] = u(\mathbb{E}[X])$$

which completes the proof $\square$

Notice that, since for a concave function $v(x)$, its negative $-v(x)$ is convex, Jensen's Inequality also implies that for a concave function $v(\cdot)$,

$$\mathbb{E}[v(X)] \leq v(\mathbb{E}[X])$$

8
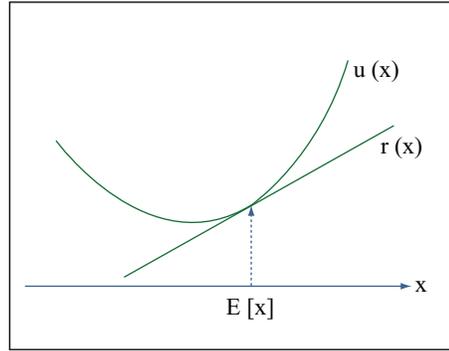
Image by MIT OpenCourseWare.

Figure 4: $r(x)$ is always less than $u(x)$

Example 9 *(Risk Aversion): Suppose you buy a laptop for 1,200 dollars which comes with a limited warranty for the first year. During that first year, there is a probability $p = 10\%$ that you spill a cup of coffee over the laptop (or have a similar accident, which is all your own fault) and have to replace the entire motherboard, which will cost 1100 dollars. This repair is not covered by the limited warranty, but you may purchase an extended service plan for 115 dollars. Should you buy this additional "insurance"? Without the additional insurance, we can think of the total cost of your laptop as a random variable which takes values $X = 1,200$ with probability $1 - p$, and $X = 1,200 + 1,100 = 2,300$ with probability p (there are different ways of setting up this problem, but let's keep things simple for now). With the extended service plan, your laptop will cost you $Y = 1,200 + 115 = 1,315$ dollars for sure.*

If you only care about the expected value of the laptop, then $\mathbb{E}[X] = 2,300p + 1,200(1-p) = 1,200 + 1,100p$. This is greater than $\mathbb{E}[Y] = 1,315$ if $p \geq 10.45\%$. But since we said that $p = 10\%$, would it still be a good idea to buy the service plan? - Economists typically assume that when people take decisions under uncertainty, they do not care about the expected amount of money W they can spend, but that they experience a utility $u(W)$ over a dollar amount, for which the additional value of an additional dollar increases in the total amount consumed. That means that we assume that $u(\cdot)$ is a concave function in costs, say

$$u(c) = \sqrt{4,800 - c}$$

where I assume that our initial wealth is $4,800$, and we can spend $4,800 - C$, where C is the total "cost" of the laptop. Then the expected utility from not having the service plan is

$$\mathbb{E}[u(C_1)] = 0.9\sqrt{4,800 - 1,200} + 0.1\sqrt{4,800 - 2,300} = 0.9 \cdot 60 + 0.1 \cdot 50 = 59$$

whereas with the insurance plan,

$$\mathbb{E}[u(C_2)] = \sqrt{4,800 - 1,315} > \sqrt{3,481} = 59$$

In fact, you would be willing to spend up to $4,800 - 3,481 = 119$ dollars for the insurance, whereas the expected additional cost is only $1,100p = 110$ dollars. This 9-dollar difference in what we are willing to pay for the insurance is referred to as the risk-premium *comes from the fact that $u(\cdot)$ is concave. By Jensen's inequality, this risk-premium is positive if $u(\cdot)$ is concave, and we say that this type of preferences exhibits* risk aversion.

Example 10 *The following example is known as the* St. Petersburg Paradox, *and it gives an example of a random variable which doesn't have a finite expectation.*
We are offered the following gamble: suppose a fair coin is tossed over and over until the first head

9

appears. You get 2 dollars when heads appears on the first flip, 2^2 dollars if it appears on the second, and, more generally, 2^k dollars if the first head appears on the xth flip.

How much would you pay to play this game? - in principle, you should be willing to pay your expected winnings, so let's do the calculation: the probability that exactly x flips are required equals

$$f_X(x) = P(x - 1 \text{ tails}, 1\text{heads}) = \frac{1}{2^{x-1}} \frac{1}{2} = 2^{-x}$$

Therefore, expected winnings Y can be calculated as

$$\mathbb{E}[Y] = \sum_{x=1}^{\infty} 2^x f_X(x) = \sum_{x=1}^{\infty} \left(\frac{2}{2}\right)^x = \sum_{x=1}^{\infty} 1 = \infty$$

Therefore, there is no upper bound on expected winnings.

Does this mean that we'd see people willing to pay an infinite amount for this type of bet? - certainly not: typically people would offer at most around 25 dollars to play the game. This paradox can be resolved in different ways:

- *people do not actually care about the expected amount of money, but their valuation of money decreases in the total amount they already have, i.e. people maximize some concave function $u(\cdot)$ of winnings, as in the previous example.*

- *with very small probabilities, the amount you win is extremely high - i.e. in the trillions, quadrillions etc. of dollars, and we would not believe that in this case, our opponent could live up to his promise, so in fact there would be some ceiling as to how much we could at best hope to win from this bet.*

In order to make the link back to Jensen's inequality, let's also calculate the expected number of coin flips in the game for a coin which comes up heads with probability $p = \frac{1}{a}$:

$$\mathbb{E}[X] = \sum_{x=1}^{\infty} x a^{-x} = \frac{1}{a} \sum_{x=1}^{\infty} x \left(\frac{1}{a}\right)^{x-1} =: \frac{1}{a} G'\left(\frac{1}{a}\right)$$

where $G'(a)$ is, as you can easily check, the first derivative with respect to $\frac{1}{a}$ of

$$G\left(\frac{1}{a}\right) = \sum_{x=1}^{\infty} \left(\frac{1}{a}\right)^x = \frac{1}{1 - \frac{1}{a}}$$

Therefore, by taking the derivative of the new expression for $G\left(\frac{1}{a}\right)$, we get that

$$\mathbb{E}[X] = \frac{1}{a} G'\left(\frac{1}{a}\right) = \frac{1}{a} \frac{1}{\left(1 - \frac{1}{a}\right)^2}$$

and since in our example, $\frac{1}{1} = \frac{1}{2}$, the expected number of flips is 2.

Hence, the 25 dollars many people are still willing to bet are far more than $2^{\mathbb{E}[X]} = 2^2 = 4$ dollars you'd get from the average number of flips. The explanation for this is once more Jensen's inequality, and the fact that $u(x) = 2^x$ is an (extremely) convex function of x.

Example 11 *Suppose you can choose between two assets: the first is the stock of an obscure online start-up which makes it possible for people to search web sites for free. It is very risky in that with probability 90% dividends are constant at $e^{0t} = 1$, and with 10% probability, the company's name is Google, and the*

dividends it pays at any instant in time t develop according to $e^{0.1T}$, i.e. there is a random growth rate G_1 which takes values 0% with probability 90%, and 10% with probability 10%, respectively. Alternatively, you can hold a government bond which will $e^{0.02t}$ interest at every instant t in the future, i.e. $G_2 = 2\%$ for sure.

You value one dollar you receive t periods from now as much as $e^{-0.15t}$ dollars you can have right now, but invest in one of the two assets anyway. I.e. in general you value an asset whose returns grow at rate g for sure as

$$V(g) = \int_0^\infty e^{(g-r)t} dt = \frac{0-1}{g-r} = \frac{1}{r-g}$$

The expected growth rate of dividends for the risky stock is

$$\mathbb{E}[G_1] = 0.9 \cdot 0\% + 0.1 \cdot 10\% = 1\% < 2\% = \mathbb{E}[G_2]$$

However, your valuation for the stock is

$$\mathbb{E}[V(G_1)] = 0.9 \frac{1}{0.15 - 0} + 0.1 \frac{1}{0.15 - 0.1} = \frac{90}{15} + \frac{10}{5} = \frac{40}{5} = 8$$

whereas you value the bond as

$$\mathbb{E}[V(G_2)] = \frac{1}{0.15 - 0.02} = \frac{100}{13} = \frac{200}{26} < \frac{200}{25} = \mathbb{E}[V(G_1)]$$

Intuitively, uncertainty over the growth rates means that even though 90% of the start-ups don't develop (or even fail), the 10% which are successful do so spectacularly as to compensate for the investments which turn bad. Formally, the function $V(g)$ is convex in growth rates, so that by Jensen's inequality investors should value risk in growth rates *- though typically not in* levels.