14.30 Introduction to Statistical Methods in Economics
Spring 2009

14.30 Introduction to Statistical Methods in Economics
Lecture Notes 12

Konrad Menzel

March 19, 2009

1 Properties of Medians and Percentiles

We defined the median of a random variable via

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

When X is discrete or has point masses that generate jumps in the c.d.f., this definition may not be useful, so in the more general case, we define the median as

$$\text{median}(X) := \min\left\{m \in \mathbb{R} : P(X \leq m) \geq \frac{1}{2}\right\}$$

The change with respect to the narrower definition is that if the c.d.f. has a discontinuity which makes it "leap" over the value $\frac{1}{2}$, just locate the median at the point of that discontinuity. We can also define other *percentiles* of the distribution of X:

Definition 1 *For a random variable X, the α quantile is given by*

$$q(X, \alpha) := \min\left\{q \in \mathbb{R} : P(X \leq q \geq \alpha\right\}$$

We also call $q(X, p/100)$ the pth percentile.

Note that following this definition, the median corresponds to the 50th percentile. Other frequently used quantiles are deciles ($p = 10, 20, 30, \ldots, 90$) and quartiles ($p = 25, 50, 75$).

Now we won't spend as much time on properties of quantiles as we did for expectations, but I'd just like to point out two important ways in which the median behaves very differently from the expectation:

For one, we saw from Jensen's Inequality that for a function $u(X)$, the expectation $\mathbb{E}[u(X)]$ depends a lot on the curvature of $u(x)$ in the regions where the probability mass of X lies. Generally, the median median$(u(X))$ will also be different from $u(\text{median}(X))$, but there is a notable exception:

Proposition 1 *Suppose $u(x)$ is* strictly increasing *in the support of X. Then*

$$\text{median}(u(X)) = u(\text{median}(X))$$

Proof: The median of X satisfies $P(X < \text{median}(X)) = \frac{1}{2}$. Since $u(x)$ is strictly increasing, the event $X < m$ is identical to the event $u(X) < u(m)$ for any fixed number m. Therefore, $P(u(X) < u(\text{median}(X)) = P(X < \text{median}(X)) = \frac{1}{2}$, so that $u(\text{median}(X))$ is indeed the median of $u(X)$ $\square$

Intuitively, the median is based only on *ordinal* properties of the random variable which are preserved by any strictly increasing transformation.

We also saw that expectations were linear in the sense that the expectation of a linear function of multiple random variables was shown to be equal to that same linear function of the expectations. For medians this is no longer true as the following example shows:

Example 1 *Suppose X_1 and X_2 are discrete random variables which are from the same marginal distribution*

$$f_X(x) = \begin{cases} 0.6 & \text{if } x = 0 \\ 0.4 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

and independent of each other. Then $Y = X_1 + X_2$ can take the values 0, 1, and 2, and has p.d.f.

$$f_Y(y) = \begin{cases} 0.36 & \text{if } x = 0 \\ 0.48 & \text{if } y = 1 \\ 0.16 & \text{if } y = 2 \\ 0 & \text{otherwise} \end{cases}$$

The median of X_1 and X_2 is zero, however $\text{median}(Y) = 1 \neq 0 + 0 = \text{median}(X_1) + \text{median}(X_2)$.

More generally, the quantiles of averages may also differ from averages of quantiles. The following example gives another very practical interpretation of this insight (thanks to Aleksandr Tamarkin for providing the following numerical example):

Example 2 *Say you are taking the GRE, a standardized test which consists of three components, verbal X_1, analytic X_2, and quantitative X_3. Your score is above the 90th percentile for each section of the test. Does this mean that you are also above the 90th percentile for the overall score? The general answer is no.*

Suppose, including yourself, there are 100 test takers, and the distribution of scores is as follows: 84 test takers don't get a single point in any section, you got 250 points in each of the three sections, and there are three other types of test takers, each with somewhat savant-like insular abilities in only one of the three areas. More specifically, 5 people are extremely verbally gifted and score 800 on the verbal part, but 0 everywhere else, another 5 get 800 on the analytic part, and yet another 5 get 800 on the quantitative part, but zero in all other sections. In sum, the joint distribution of scores is (note that this is very far from the typical distribution of GRE scores...)

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = \begin{cases} 0.84 & (x_1,x_2,x_3) = (0,0,0) \\ 0.01 & (x_1,x_2,x_3) = (250,250,250) \text{ (you)} \\ 0.05 & (x_1,x_2,x_3) = (800,0,0) \\ 0.05 & (x_1,x_2,x_3) = (0,800,0) \\ 0.05 & (x_1,x_2,x_3) = (0,0,800) \\ 0 & \text{otherwise} \end{cases}$$

So you are at least at the 95th percentile for each part, but for 15 other test takers, the total score is 800, whereas yours is only 750, so you are only at the 85th percentile with respect to the overall score across the three sections.

2 Variance

The *variance* is a measure of the dispersion of a random variable.

Definition 2 *The* variance *of a random variable X is given by*

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

Sometimes we also denote the variance by $\sigma^2(X) = \mathrm{Var}(X)$.

Property 1 $\mathrm{Var}(X) = 0$ *if and only if $P(X = c) = 1$ for some constant c.*

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

$$\mathrm{Var}Y = a^2\mathrm{Var}(X)$$

PROOF: Again, we'll only look at the continuous case. Using our previous results on expectations

$$\begin{aligned}
\mathrm{Var}(Y) &= \int_{-\infty}^{\infty}(aX + b - \mathbb{E}[aX + b])^2 f_X(x)dx = \int_{-\infty}^{\infty}(ax - a\mathbb{E}[X])^2 f_X(x)dx \\
&= a^2\int_{-\infty}^{\infty}(x - \mathbb{E}[X])^2 f_X(x)dx = a^2\mathrm{Var}(X)
\end{aligned}$$

It is often convenient for the measure of dispersion to have the same units as the random variable. However, this last result implies that the unit of $\mathrm{Var}(X)$ will be the square of the unit of X. Therefore, we often use the *standard deviation* $\sigma(X)$ instead,

$$\sigma(X) := \sqrt{\mathrm{Var}(X)}$$

Property 3

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

PROOF:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\right] \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2
\end{aligned}$$

Property 4 *If $Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n + b$ and $X_1, \ldots, X_n$ are* independent*, then*

$$\mathrm{Var}(Y) = a_1^2\mathrm{Var}(X_1) + a_2^2\mathrm{Var}(X_2) + \ldots + a_n^2\mathrm{Var}(X_n)$$

Example 3 *Suppose X is a discrete random variable with p.d.f.*

$$f_X(x) = \begin{cases} \frac{1}{5} & \text{if } x \in \{-2, 0, 1, 3, 4\} \\ 0 & \text{otherwise} \end{cases}$$

If we define $Y = 4X - 7$, what is the variance of Y?

$$\mathrm{Var}(Y) = 4^2\mathrm{Var}(X) = 16\left(\mathbb{E}[X^2] - \mathbb{E}[X]^2\right)$$

We can now calculate

$$\mathbb{E}[X] = \frac{1}{5}(-2 + 0 + 1 + 3 + 4) = \frac{6}{5}$$

and

$$\mathbb{E}[X^2] = \frac{1}{5}((-2)^2 + 0^2 + 1^2 + 3^2 + 4^2) = \frac{30}{5} = 6$$

Therefore,

$$\mathrm{Var}(Y) = 16\left[6 - \left(\frac{6}{5}\right)^2\right] = 16\frac{150-36}{25} = \frac{1824}{25} \approx 73$$

Example 4 *Suppose $Y \sim B(n,p)$. Since Y can be written as the sum of outcomes from n* independent *trials,*

$$Y = X_1 + \ldots + X_n, \text{where } X_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

we can calculate

$$\mathrm{Var}(Y) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)$$

So what is the variance of X_i? Clearly,

$$\mathbb{E}[X_i] = p$$

also,

$$\mathbb{E}[X_i^2] = 1 \cdot p + 0 \cdot (1-p) = p$$

Therefore, by property 3

$$\mathrm{Var}(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = p - p^2 = p(1-p)$$

Therefore

$$\mathrm{Var}(Y) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = np(1-p)$$

Since the variance is an expectation, we can directly apply results on expectations of functions of random variables directly to the *variance* of a function of random variables: if $Y = r(X)$,

$$\mathrm{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = \mathbb{E}[r(X)^2] - \mathbb{E}[r(X)]^2 = \int_{-\infty}^{\infty} r(t)^2 f_X(t)dt - \left[\int_{-\infty}^{\infty} r(t)dt\right]^2$$

2.1 Higher-Order Moments

We saw that the expected value is a measure of the *location* of a distribution, whereas the variance measures its *dispersion*. We may look at other *moments* of the random variable in order to characterize its distribution, e.g. whether it's symmetric, has fat tails etc.

Definition 3 *The* r-th moment *of X is given by*

$$\mu_r' = \mathbb{E}[X^r]$$

and the r-th central *moment is defined as*

$$\mu_r = \mathbb{E}[(X - \mathbb{E}[X])^r]$$

The expectation is therefore also referred to as the *first moment* of the distribution of X, and the variance as its *second central moment*.
Other frequently used characteristics of a distribution are

$$\mu_3 = \mathbb{E}[(X - \mathbb{E}[X])^3]$$

which is called the *skewness* of the distribution of X, and

$$\mu_4 = \mathbb{E}[(X - \mathbb{E}[X])^4]$$

the *kurtosis* of X. A high kurtosis corresponds to the distribution having a lot of probability mass in the tails.