But since variances are always non-negative,

$$V_X\left[E(Y \mid X)\right] \ge 0 \;,\; \text{so}$$

$$\underbrace{E_X\left[V(Y \mid X)\right] + V_X\left[E(Y \mid X)\right]}_{\substack{V(Y) \\ \text{MSE of } \hat{Y}_{\text{no } X}}} \ge \underbrace{E_X\left[V(Y \mid X)\right]}_{\substack{\text{"E(MSE)"} \\ \text{of } \hat{Y}_X}}$$

Thus you always expect your predictive accuracy to get better (or at least stay the same) when you use $E(Y \mid X)$ to predict Y.

Another complete switch in subject!

Utility

$\boxed{\text{Q:}}$ How to take action sensibly when the consequences are uncertain?

A: There is a theory of optimal action under uncertainty; it's called Bayesian decision theory — a concept called utility is central to this theory. The theory takes its simplest form when comparing ~~[illegible]~~ gambles

Example: X has discrete PMF $f_X(x) = \begin{cases} \frac{1}{2} & x = -\$350 \\ \frac{1}{2} & x = +\$500 \\ 0 & \text{else} \end{cases}$

Suppose X = your net gain from gamble (A), Y has discrete PMF $f_Y(y) = \begin{cases} \frac{1}{3} & y = \$40 \\ \frac{1}{3} & y = \$50 \\ \frac{1}{3} & y = \$60 \\ 0 & \text{else} \end{cases}$

and Y = your net gain from gamble (B).

Turns out that $E(X) = \$+75$, $E(Y) = \$+50$, so is (A) automatically better than (B)?

Note that with B you're guaranteed to win at least 840, while A has no such guarantee; is A still automatically better **for you** than B? A risk-averse person would grab B quickly; a risk-seeking person would probably pick A.

Evidently something more than just computing $E(X)$, $E(Y)$ is going on.

DS def. of utility function

Your utility function $U(x)$ is that function which assigns to each possible net gain $-\infty < x < \infty$ a real # $U(x)$ representing the value to you of gaining x.

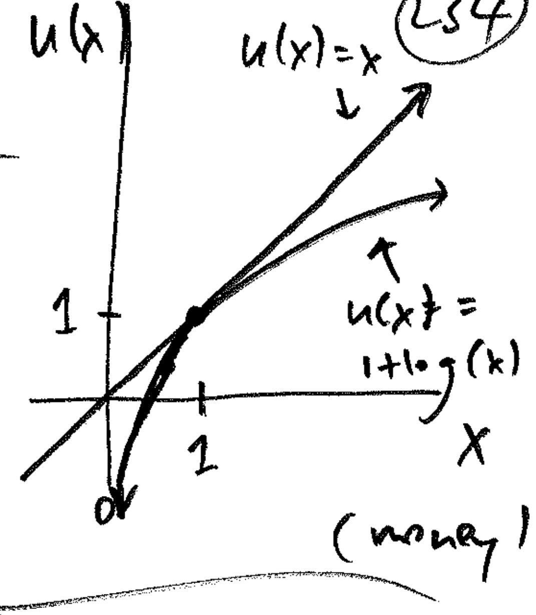
Q: If x is money, why not just use $u(x) = x$? (utility linear in money)

A: Lovely, subtle answer first supplied by Daniel Bernoulli (1700 - 1782), (Swiss mathematician) related to Jacob Bernoulli (1654 - 1705), for whom the Bernoulli distribution was named.

Daniel B: If your entire net worth is (say) \$10, then the value to you of a new \$1 is much greater than if your entire net worth is (say) \$1,000,000; thus the utility of money is sublinear (meaning that it doesn't grow with x as fast as $f(x) = x$ does)

Daniel B proposed one particular sublinear function for utility, namely $u(x) = 1 + \log(x)$ (for $x > 0$)

(Daniel B also invented the word <u>utility</u>) (Although the idea goes back at least to Aristotle (384-322 BCE))



<u>Definition</u>

(Principle of Expected Utility maximization)

You are said to choose between gambles by <u>maximizing</u> <u>expected utility</u> (MEU) if, with $U(x)$ your utility function, (a) you prefer gamble X to gamble Y if $E[U(X)] > E[U(Y)]$ and (b) you're indifferent between X and Y if $E[U(X)] = E[U(Y)]$.

MEU first explored in depth by British

{mathematician, philosopher, economist} Frank Ramsey (1903 - 1930), who died at age 26 of liver failure. (hepatitis)

Theorem (von Neumann - Morgenstern (1947)) under 4 reasonable axioms, MEU is the best you can do.

John von Neumann (1903 - 1957) Hungarian - American {mathematician, physicist, computer scientist, ⋮} died at 53 of cancer

Oskar Morgenstern (1902 - 1977) German-American economist

Simple example: Suppose you bought a single \$2 ticket in the Powerball lottery examined in Take-Home Test 1 problem 2: the drawing on 30 Jul 2016 for which the Grand prize was \$487 million. Let X be the unknown amount you will win (thinking about X before the drawing).

Match	x	$P(X=x)$	$x \cdot P(X=x)$ 236
5W, 1R	\$487,000,000	$\frac{1}{292,201,338}$	\$1.667
5W, 0R	\$1,000,000	$\frac{1}{11,688,053.52}$	0.086
4W, 1R	\$50,000	$\frac{1}{913,129.18}$	0.055
4W, 0R	\$100	1/36,525.17	0.003
3W, 1R	\$100	1/14,494.11	0.007
3W, 0R	\$7	1/579.76	0.012
2W, 1R	\$7	$\frac{1}{701.33}$	0.010
1W, 1R	\$4	$\frac{1}{91.98}$	0.043
0W, 1R	\$4	$\frac{1}{38.32}$	0.104
			$\boxed{\$1.99 \ (!)}$

X has 9 possible values x (discrete),

so $E(X) = \sum_{\substack{\text{all} \\ 9 \text{ possibilities}}} x \cdot P(X = x) = \$1.99$

Q: Before the drawing, someone offers you $\$x_0$ for your ticket; should you sell?

A: With $U(x)$ as your utility function, your expected gain if you keep the ticket is $E[U(X)]$; if for you $U(x) = x$ (utility $\hat{=}$ money) then $E(U(X)) = \$1.99$

Action 1 (sell): you gain $\$x_0$ for sure

Action 2 (keep): your expected utility is $E[U(X)]$

Under MEU you should sell if $\boxed{U(x_0) > E[U(X)]}$

If $U(x) = x$ for you then your optimal action is (sell if offered more than $1.99).

Related but different problem

On the 13 Jan 2016 drawing the Powerball jackpot was $1.6 billion!

(238)

X = your winnings

X uncertain before the drawing

Redo calculation on p. (236): $E(X)$ is now \$5.80 on a \$2 ticket

new 1st row in table is $\frac{1{,}600{,}000{,}000}{292{,}201{,}338} \doteq 5.476$

Q: If $u(x) = x$ for you, is it rational to sell all your assets & buy as many lottery tickets as possible? (*)

A: Yes, but that's a silly utility function; to be realistic you'd have to subtract from x the monetary value (cost) to you of the disruption of your life that would ensue with action

(*) A catalog of useful distributions

(DS ch.5) Case 1: Discrete — Bernoulli

$X \sim \text{Bernoulli}(p)$, $0 < p < 1$, if

$$f_X(x) = p^x (1-p)^{1-x} \, \overbrace{I_{\{0,1\}}(x)}^{\text{support}(X)}$$

$$= \begin{cases} p & \text{for } x = 1 \\ 1-p & 0 \\ 0 & \text{else} \end{cases}$$

$E(X) = p$

$V(X) = p(1-p)$

$\psi_X(t) = p e^t + (1-p)$ for all $-\infty < t < \infty$

$SD(X) = \sqrt{p(1-p)}$

Def If the X_i in $X_1, X_2, \dots$ are IID Bernoulli (p), then $(X_1, X_2, \dots)$ are called <u>Bernoulli trials</u> with <u>parameter</u> p; if the sequence $(X_1, X_2, \dots)$ is infinite this defines a <u>Bernoulli (stochastic) process</u>.

Binomial $X \sim$ Binomial (n, p) (i.e., X follows the Binomial distribution with parameters n $\left(\begin{matrix}\text{positive}\\\text{integer}\end{matrix}\right)$ and $0 < p < 1$)

$$\leftrightarrow \quad f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \, \overbrace{I_{\{0, 1, \dots, n\}}}^{\text{support}(X)}(x)$$

Consequences ① $X_1, \dots, X_n \overset{\text{IID}}{\sim}$ Bernoulli(p)

$$\rightarrow \quad X = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p)$$

$X \sim \text{Binomial}(n, p)$ | $E(X) = n \cdot p$ | $V(X) = np(1-p)$

$\psi_X(t) = [pe^t + (1-p)]^n$ for all $-\infty < t < \infty$

$SD(X) = \sqrt{np(1-p)}$

Case Study | Castaneda v. Partida (1977)

Grand juries in the U.S. judicial system have catchment areas: everybody 18 & over living in the judicial district for that grand jury (& a few other minor restrictions)

Hidalgo County, Texas

extreme southern border of TX with Mexico

eligible pool was 79.1% Mexican-American

$2\frac{1}{2}$ yr period at issue in Supreme Court case: 220 people called to serve on grand juries, but only 100 of them were Mexican-American

Q: Prima facie case of discrimination?

Before this $2\frac{1}{2}$ yr period, let X be your prediction of # of Mexican-Americans among the 220 people

If no discrimination,

$X \sim \text{Binomial}(220, 0.791)$
$(X \mid T_1)$

T_1 = theory 1 = no discrimination

$$E(X \mid T_1) = \overset{n}{(220)} \cdot \overset{p}{(0.791)} \doteq 174.0$$

$$SD(X \mid T_1) = \sqrt{n\,p\,(1-p)} \doteq 6.0$$

Q: If you were expecting 174 give or take 6, would you be surprised to see 100?

A: You'd be astonished

Frequentist statistical answer: $P(X \le 100 \mid T_1) = 8.0 \cdot 10^{-28}$ so T_1 looks ridiculous

Bayesian statistical answer: Need to compute $P(T_1 \mid X = 100)$, not the other way around (later)

Hypergeometric | A finite population has A elements of type 1 and B elements of type 2; total population size $(A+B)$.

You choose n elements at random <u>without</u> replacement from this population (i.e., you take a <u>simple random sample</u> (SRS) of size n) | Let $X = \left(\text{\# elements of type 1 in your sample}\right)$

Then (as noted in take-Home Test 1 problem 2) X follows the <u>hypergeometric</u> distribution with parameters (A, B, n). As we saw in that problem, the PMF of X is

$$f_X(x \mid A, B, n) = \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}} \, I\big[\underbrace{\max\{0, n-B\} \le x \le \min\{n, A\}}_{\text{support}(X)}\big]$$

for (A, B, n) non-negative integers with $n \le A+B$

Consequences

① $E(X) = n \cdot \dfrac{A}{A+B}$

② $V(X) = n\left(\dfrac{A}{A+B}\right)\left(\dfrac{B}{A+B}\right)\left(\dfrac{A+B-n}{A+B-1}\right)$

Note that if your sampling had been with replacement (i.e., you take an IID sample), X would have been Binomial with the same value of n and $p = \dfrac{A}{A+B}$; in that case $E(X) = np = n\dfrac{A}{A+B}$ and

$V(X) = np(1-p) = n\left(\dfrac{A}{A+B}\right)\left(\dfrac{B}{A+B}\right)$. (compare)

If you let $T = (A+B)$ be the total # of elements in the population,

sampling method	mean	variance
with repl. (IID)	$n\left(\frac{A}{A+B}\right)$	$n\left(\frac{A}{A+B}\right)\left(\frac{B}{A+B}\right)$
without repl. (SRS)	$n\left(\frac{A}{A+B}\right)$	$n\left(\frac{A}{A+B}\right)\left(\frac{B}{A+B}\right)\left(\frac{T-n}{T-1}\right)$

$0 \leq \alpha = \frac{T-n}{T-1} \leq 1$ is called the finite population correction

3 special cases worth considering

(a) $(n=1)$ $\alpha = 1 \leftrightarrow$ SRS = IID with only 1 element sampled

(b) $(n=T)$ $\alpha = 0 \leftrightarrow$ If you exhaust the entire population ~~on~~ with SRS, you have no uncertainty left.

(c) (n fixed, $T \uparrow$) $\alpha \uparrow 1$ $\leftrightarrow$ with a small sample from a large population,

SRS $\doteq$ IID

Poisson ($\lambda > 0$) $X \sim \text{Poisson}(\lambda)$

$\leftrightarrow$ X has PMF $f_X(x) = \dfrac{\lambda^x e^{-\lambda}}{x!} \underbrace{I_{\{0,1,\ldots\}}(x)}_{\text{support of } X}$

$\boxed{E(X) = \lambda}$

$\boxed{V(X) = \lambda}$

$\Big\}$ thus for the Poisson dist.

$$\frac{V(X)}{E(X)} = 1$$

$\boxed{\text{Def.}}$ If $E(X)$ and $V(X)$ both exist and $E(X) \neq 0$, $\dfrac{V(X)}{E(X)}$ is called the variance-to-mean ratio (VTMR)

$\psi_X(t) = e^{\lambda(e^t - 1)}$

$-\infty < t < \infty$

The Poisson ~~[scribble]~~ can be unrealistic as a consequence of its VTMR of 1, $\rightarrow$ because many rvs that represent counts of occurrences of events in time intervals of fixed length have VTMR > 1.

The Poisson & Binomial distributions both count the numbers of "successes" in a process unfolding in time, so it should not be surprising to find out that these 2 dist. are related:

---

when $\left(\begin{array}{l} n \text{ is large} \\ p \text{ is close to } 0 \end{array} \right)$, $\text{Binomial}(n, p) \doteq \text{Poisson}(n \cdot p)$

Theorem n positive integer, $0 < p < 1$, $X \sim \text{Binomial}(n, p)$

$\lambda > 0$, $Y \sim \text{Poisson}(\lambda)$ / Choose any sequence