Then the rv's $(S_i | \Theta)$ are IID Bernoulli$(\theta)$

and the rv $S = \sum_{i=1}^{n} S_i$ has a conditional binomial dist: $(S | \Theta) \sim \text{Binomial}(n_T, \theta)$

It's meaningful to talk about the conditional expectation rv: $E(S | \Theta) = n_T \theta$ (a linear function of $\Theta$),

and - via Bayes Theorem - it's even more meaningful to talk about the conditional expectation rv: $E(\Theta | S)$ (more about this later)

and the constant $E(\Theta | S = s)$. Important consequence of the def. of conditional expectation

Remember the Law of Total Prob?!
Continuous version of LnP

for which all named densities exist

\[ f_Z(y) = \int_{-\infty}^{\infty} f_X(x) \cdot f_{Z|X}(y|x) \, dx \]

Earlier we agreed that, by definition,

\[ E(Z|X) = \int_{-\infty}^{\infty} y \cdot f_{Z|X}(y|x) \, dy \]

So until the following slightly magical,

\[ E(Z) = \int_{-\infty}^{\infty} y \left( \int_{-\infty}^{\infty} f_X(x) \cdot f_{Z|X}(y|x) \, dx \right) \, dy \]

if ok to interchange order of integration.
This is referred to as part 2 of the double expectation theorem. Strictly, it is of the form

\[ \int_{-\infty}^{\infty} E(\hat{Y}) \, dE(x) \]

Recall that for a continuous random variable, \( E(Y) = \int_{-\infty}^{\infty} y \, f_Y(y) \, dy \).

So \( \hat{Y} \) is just the weighted average of \( E(Y(x)) \) with \( f_X(x) \) as the weights.

\[ \hat{Y} = \int_{-\infty}^{\infty} E(Y(x)) \, f_X(x) \, dx \]

\( \hat{Y} \) is not the mean of \( Y \), as some call it, that is \( E(Y) \).

Instead, it is the weighted average, etc.
I need to postpone examples of these conditional expectation calculations until we've covered more standard distributions.  

If $Y$ is a r.v. such that $f_{Y|X}(y|x)$ exists — it makes sense to speak not only of $E(Y|X)$, the mean of $f_{Y|X}(y|x)$, but also of the variance of that dist.

$$\text{Def } V(Y|X) = \mathbb{E}\left\{ (Y - E(Y|X))^2 | X \right\}$$

is called the conditional variance of $Y$ given $X = x$, and the r.v. $V(Y|X)$ is just $g(X)$, the conditional variance of $Y$ given $X$.  

\[ g(x) \]
followingly express the exist, \[ E(\xi) = E(\xi) \]

\[ E[\xi_{\text{predict}}] = E(\xi) \] the conditional

\[ \text{MSE} = E[(\xi - \hat{\xi})^2] \]

\[ \hat{\xi} = \hat{\theta}(\xi) \text{ to predict } \hat{\xi} \text{ from } \xi \to \]

\[ \hat{\theta} = \theta(\xi) \text{ at minimize} \]

\[ \text{for all } (\hat{\xi}, \xi) \text{ related to } \]

\[ \text{inflation} \]

\[ \text{inflation} \]
Imagine a 2-part game.

**Stage 1** Predict \( \hat{Y} \) without knowing \( X \).

Well, if you buy into MSE as your measure of "goodness" of a prediction, we know that you should predict \( \hat{Y} = \mu_X = E(Y) \).

and your result is \( \text{MSE} \) will be

\[
E[(Y - \hat{Y})^2] = \sqrt{(Y)} = \sigma^2_X.
\]

**Stage 2** Observe \( X \), now predict \( \hat{Y} \).

Let's say \( \delta = x^* \).

Then we know the MSE-optimal prediction is \( \hat{Y} = E(Y | X = x^*) \).
and your resulting MSE will be

\[ E\{[\hat{Y} - E(Y | \hat{X} = x^*)]^2\} = V(Y | x^*) \]

(229)

From the vantage point of someone thinking about stage 2 before it happens, \( X \) is not yet known, so the expected value of \( \hat{X} \), namely \( E_{\hat{X}}[V(Y | \hat{X})] \), is the best you can do to guess at how good the stage 2 prediction will be. The second part of the double expectation theorem says

\[ V(Y) = E_{\hat{X}}[V(Y | \hat{X})] + V_{\hat{X}}[E(Y | \hat{X})] \]

\[ \overline{\text{MSE of } \hat{Y}_{\text{no}X}} \]

\[ \overline{\text{"E(MSE)" of } \hat{Y}_{\text{no}X}} \]

\[ \overline{\hat{Y}_{\text{no}X} = E(Y | \hat{X})} \]
But since variances are always non-negative,

\[ V_X \left[ E(Y | X) \right] \geq 0, \quad \therefore \]

\[ E_X \left[ V(Y | X) \right] + E_X \left[ V(E(Y | X)) \right] \geq E_X \left[ V(Y | X) \right] \]

\[ \sqrt{\hat{V}(\hat{Y})} \geq \text{MSE of } \hat{Y}_{\text{no}\ X} \]

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

**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 gambles.

Example: \( X \) has PF \( f_X(x) = \begin{cases} \frac{1}{2} & x = -350 \\ \frac{1}{2} & x = 850 \\ 0 & \text{else} \end{cases} \)

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

and \( Y \) is your net gain from gamble \( \Box \). Turned out that \( \text{So is } \Box \) automatically better.

\( E(X) = -75 \), \( E(Y) = 50 \) from \( \Box \)
value to you of gaining x.

Define your utility function, u(x), which assigns a utility value to each possible net gain.

Subjectively, E[u(x)] is going on.

Evidently, something more than just

risk-seeking person would, 0.21

win at least 84, while no

such guarantee is, 0.0

A risk-averse

than [for you] plan B?

θ
If $x$ is money, why not just use $u(x) = x$? 

A: lovely, subtle answer first (utility is money)

Supplied by Daniel Bernoulli (1700 - 1782), a 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 \( x > 0 \))

(Although the idea goes back at least to Aristotle (384-322 BCE) Definition)

Principle of Expected Utility Maximization

You are said to choose between gambles \( X \) and \( Y \) by maximizing expected utility if, with \( u(x) \) your utility function,

\( \circ \) you prefer gamble \( X \) to gamble \( Y \) if \( E[u(X)] > E[u(Y)] \)

\( \circ \) you're indifferent between \( X \) and \( Y \) if \( E[u(X)] = E[u(Y)] \)

\( \circ \) you prefer gamble \( Y \) to gamble \( X \) if \( E[u(X)] < E[u(Y)] \)
The expected (von Neumann-Morgenstern, 1947) under 4 reasonable axioms, MEU is the best you can do.

Simple example: You bought a single $2 ticket in the powerball lottery. Examined in homework 1 problem 2: the drawing on 30 Jul 2018 for which the grand prize was $487 million. Let I be the amount you will win (think of cost I before the drawing).
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 = 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 $U(x_0) \geq E[U(x)]$

If $U(x) = x$ for you then your optimal action is (sell if offered more than $1.99$).
<table>
<thead>
<tr>
<th>Match</th>
<th>x</th>
<th>$P(X=x)$</th>
<th>$x \cdot P(X=x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5w, 1R</td>
<td>$487,000,000$</td>
<td>$\frac{1}{292,201,338}$</td>
<td>$1.667$</td>
</tr>
<tr>
<td>5w, 0R</td>
<td>$1,000,000$</td>
<td>$\frac{1}{1,585,653.52}$</td>
<td>$0.080$</td>
</tr>
<tr>
<td>4w, 1R</td>
<td>$850,000$</td>
<td>$\frac{1}{943,129.16}$</td>
<td>$0.055$</td>
</tr>
<tr>
<td>4w, 0R</td>
<td>$100$</td>
<td>$\frac{1}{136,572.51}$</td>
<td>$0.007$</td>
</tr>
<tr>
<td>3w, 1R</td>
<td>$$100$</td>
<td>$\frac{1}{14,494.11}$</td>
<td>$0.07$</td>
</tr>
<tr>
<td>3w, 0R</td>
<td>$7$</td>
<td>$\frac{1}{579.76}$</td>
<td>$0.012$</td>
</tr>
<tr>
<td>2w, 1R</td>
<td>$7$</td>
<td>$\frac{1}{201.33}$</td>
<td>$0.010$</td>
</tr>
<tr>
<td>1w, 1R</td>
<td>$84$</td>
<td>$\frac{1}{91.58}$</td>
<td>$0.043$</td>
</tr>
<tr>
<td>0w, 1R</td>
<td>$4$</td>
<td>$\frac{1}{38.32}$</td>
<td>$0.104$</td>
</tr>
</tbody>
</table>

$\sum$ has 9 possible values $x$ (discrete),

$E(X) = \sum_{x} x \cdot P(X=x)$ = $81.99$
Related but on 13 Jan 2016 knowing the Powerball jackpot was $1.6 billion

If your winnings are uncertain before the draw,

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

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

new first row in table is $\frac{1,600,000,000}{292,201,330} = 85.476$
A catalog of useful distributions

(Och. 5) Case 1: Discrete

Bernoulli

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

\[ f_X(x) = p^x (1-p)^{1-x} \quad \text{I}_{[0,1]}(x) \]

\[ E(X) = p \]

\[ \mu_X(t) = pe^t + (1-p) \quad \text{for all } -\infty < t < \infty \]

\[ \sigma^2(X) = p(1-p) \]

\[ \text{SD}(X) = \sqrt{p(1-p)} \]
Def: If the $X_i$ in $X_1, X_2, \ldots$ are independent Binomial ($p$), then $(X_1, X_2, \ldots)$ are called Bernoulli trials with parameter $p$. If the sequence $(X_1, X_2, \ldots)$ is infinite, this defines a Bernoulli (stochastic) process.

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

$\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x \in \{0, 1, \ldots, n\}$

Consequences $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$

$\Rightarrow \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p)$
\( X \sim \text{Binomial}(n, p) \)  
\[ E(X) = np \quad V(X) = np(1-p) \]

\[ P(t) = \left[ pe^t + (1-p) \right]^n \quad \text{for all } -\infty < t < \infty \]

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

**Example**  
Cartoneda v. Partida (1977)

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

Hidalgo County, Texas

2½ yr period at issue in Supreme Court case: 220 people called to serve on grand juries, but only 100 of them were Mexican-American

Eligible pool was 74.1% Mexican-American

Q: Prima facie case of discrimination?
Before this 21yr period, let $X$ be your prediction of the # of Mexican-Americans among the 220 people. If no discrimination,

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

$E(X | T_1) = (220)(0.791) = 174.0$

$\mu \approx n \cdot p \approx 220 \cdot 0.791 = 174.0$

$\sigma = \sqrt{n \cdot p \cdot (1-p)} \approx 6.0$

you were expecting 174 give or take 6, would you be surprised to see 100? [A: You'd be astonished]

- Frequentist statistical answer

$P(X \leq 100 | T_1) = 8.0 \cdot 10^{-28}$

$T_1$ looks ridiculous

- Bayesian statistical answer

Need to compute $P(T_1 | X = 100)$, 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 without replacement from this population (i.e., you take a simple random sample (SRS) of size \( n \)).

Let \( X = \text{number of elements of type 1 in your sample} \). According to Homework 1, Problem 2, \( X \) follows the hypergeometric distribution with parameters \((A, B, n)\). As we saw in that problem, the PMF of \( X \) is
\[ f(x | A, B, n) = \binom{A}{n} \binom{B}{n-x} \frac{1}{\binom{A+B}{n}} I[\max\{0, n-B\} \leq x \leq \min\{n, A\}] \quad (\text{support } x) \]

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

Consequences:

1. \(E(x) = n \cdot \frac{A}{A+B}\)

2. \(\text{Var}(x) = n \left( \frac{A}{A+B} \right) \left( \frac{B}{A+B} \right) \frac{A+B-n}{A+B-1} \quad \text{Note that if your sampling has been with replacement (i.e., you take an IID sample), } x \text{ would have been binomial with the same value of } n \text{ and } p = \frac{A}{A+B} \text{ in that case } E(x) = np = n \cdot \frac{A}{A+B} \text{ and } \]

\[ \text{Var}(x) = np(1-p) = n \left( \frac{A}{A+B} \right) \left( \frac{B}{A+B} \right). \text{ (Compare)} \]
If you let \( T = (A + B) \) be the total number of elements in the population,

<table>
<thead>
<tr>
<th>Sampling Method</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>With repl. (IID)</td>
<td>( n \left( \frac{A}{A+B} \right) )</td>
<td>( n \left( \frac{A}{A+B} \right) \left( \frac{B}{A+B} \right) )</td>
</tr>
<tr>
<td>Without repl. (SRS)</td>
<td>( n \left( \frac{A}{A+B} \right) )</td>
<td>( n \left( \frac{A}{A+B} \right) \left( \frac{B}{A+B} \right) \left( \frac{T-n}{T-1} \right) )</td>
</tr>
</tbody>
</table>

\[ 0 < \frac{T-n}{T-1} \leq 1 \] is called the finite population correction. 3 special cases are worth considering:

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

2. \( n = T \)  \( \alpha = 0 \)  \( \leftrightarrow \) If you exhaust the entire population in SRS, you have no uncertainty
(c) (n fixed, \( T \uparrow \)) \[ \text{as with a} \]

small sample from a large population, \( S_n \rightarrow \text{IID} \]

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

\[ \Rightarrow X \text{ has PDF } f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!} I_{\{0, 1, \ldots\}}(x) \]

\[ E(X) = \lambda \]

\[ V(X) = \lambda \]

\[ g_X(t) = e^{\lambda(t-1)} \quad -\infty < t < \infty \]

The Poisson can be unrealistic as a consequence of its variance-to-mean ratio \( \frac{V(X)}{E(X)} = 1 \), because \( \frac{V(X)}{E(X)} \neq 0 \) if \( E(X) \) and \( V(X) \) both exist and \( E(X) \neq 0 \).
many rvs that represent counts of occurrences of events in time intervals of fixed length have \( \text{vTrR} > 1 \).

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

when \( n \) is large, \( p \) is close to 0, \( \text{Binomial} (n, p) \approx \text{Poisson} (n \cdot p) \).

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

\( \to \), \( X \sim \text{Poisson}(\lambda) \) / choose any sequence
Example

(3) Non-overlapping

(b) Arrivals in all disjoint

The time intervals

\( f(\lambda t) \theta \infty \lim_{n \to \infty} \frac{1}{n} \)

of values between 0 and 1 with

\(= \emptyset \)
organism called *cryptosporidium* that's capable of getting into the public drinking water supplies; at one stage in their life cycle they're called oocysts. They can make people sick at a concentration of only 1 oocyst per 5 liters = 1.3 gallons of water.

One problem is that it can be hard to detect these oocysts with water filtration. Suppose that, in the water supply of your city, oocysts occur according to a Poisson process with rate 2 oocysts per liter, and that the filtering system your water utility company uses can capture all the oocysts in a water sample but only has
probability $p$ of detecting each oocyst that's actually there. \((\&\text{counting events are independent})\)

Set $Y = \#\text{oocysts in } t\text{ liters of water}$, and $X_i = \{1 \text{ if oocyst } i \text{ gets counted} \}
\text{ else}\

\[X = \#\text{ counted oocysts}\] \text{ then } (X|Y=y) = \sum_{i=1}^{y} X_i\]

Under these assumptions, \((X|Y=y) \sim \text{Binomial}(Y, p)\)

Q: What's the dist. of $X$? A: By the law of total probability

\[f_X(x) = P(X=x) = \sum_{y=0}^{\infty} P(Y=y)P(X=x|Y=y)\]

for all $x = 0, 1, \ldots$

in which $P(Y=y) = (\frac{3e}{y!})^{\frac{3e}{y!}}$ for $y=0, 1, \ldots$
\[ P(X = x \mid X = y) = (1 - p)^{y-x} p^x \] Notice that if \( X = x \), \( X \geq x \) because the number of oocysts \( (X) \) has to be at least as large as the number of oocysts detected \( (X) \). After a careful calculation,

\[
\sum_{y=x}^{\infty} (1 - p)^{y-x} p^x \frac{y!}{(y-x)!} e^{-\mu t} \] \[ \mu t \]

\[ = e^{-\mu t} (\mu t)^x \frac{x!}{x!} \]

is Poisson \((\mu t)\): losing a proportion \((1-p)\) of the oocysts to faulty counting just lowers the rate of the Poisson process from \( \lambda / \text{liter} \) to \( \lambda (1-p) / \text{liter} \) (makes excellent sense).
In practice, we must cut \( p \) is small (not far from 0). \( \square \) How much water do you need to filter to achieve \( P(\text{at least 1 oocyte detected}) \geq 1 - \alpha \) for small \( \alpha \)? \( \square \) Not had to work out

\[
P(\text{at least 1 detected}) = 1 - P(\text{none detected}) = 1 - e^{-p^2 t} \geq 1 - \alpha
\]

\[t \geq -\frac{\ln \alpha}{p^2} \iff \ln \alpha \leq -p^2 t \iff \]

\[
t = -\frac{\ln \alpha}{p^2}
\]

Example \( p = 0.1, \alpha = 0.1 \)

\[
a = 0.2 \text{ liter} \quad (1 \text{ per 5 liters})
\]

minimum sickness level

\[
t \geq 230.3 \text{ liters} \quad (17\text{April 2012})
\]