Stat 265 – Unit 3 – Video Set 3 – Class Examples

Summary:

1) Bernoulli Trial: A trial resulting in a "success" with probability *p* and a "failure" with probability 1 – *p*.

2) Bernoulli(*p*) Random Variable:

- X = the number of *successes* on one *Bernoulli trial* $\sim Bernoulli(p)$.
- $p_X(x) = P(X = x) = p^x (1-p)^{1-x}, \; x = 0,1.$
- $E[X^k] = p, \; \mu_X = E[X] = p, \quad \sigma_X^2 = V[X] = p(1-p).$

$X = \begin{cases} 1, & \text{if a success} \\ 0, & \text{otherwise} \end{cases}$

3) Binomial(*n, p*) Random Variable: Independent and Identically Distributed.
- Y = the number of *successes* on n iid *Bernoulli trials* $\sim Binomial(n, p)$.
- $p_Y(y) = P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}, \; y = 0,1,2,\ldots,n.$
- $\mu_Y = E[Y] = np, \; \sigma_Y^2 = V[Y] = np(1-p).$

4) Other Properties/Notes for Binomial Random Variables:

- If $n = 1$ then $Y \sim Binomial(n = 1, p) \sim Bernoulli(p)$.
- If $X_i \overset{ind}{\sim} Bernoulli(p), \; i = 1,\ldots,n,$ then $Y = \sum_{i=1}^{n} X_i \sim Binomial(n, p)$.
- If $Y_i \overset{ind}{\sim} Binomial(n_i, p), \; i = 1,\ldots,k,$ then $Y = \sum_{i=1}^{k} Y_i \sim Binomial\left(\sum_{i=1}^{k} n_i, p\right)$.

Example A: When the health department tested private wells in a county for impurities commonly found in drinking water, it was found that 20% of the wells had neither impurity, 40% had impurity A, and 50% had impurity B. Consider a random sample of ***n*** wells. Assume all independent.

a) If they inspect ***n*** = 10 wells, what is the probability they will find at least 3 wells with both impurities?

Let Y = # of wells out of 10 with both impurities ↳ success.

$\sim \text{Binomial}(n = 10, p = P(A \cap B) = 0.10)$

$$p_Y(y) = \binom{10}{y}(0.10)^y(0.90)^{10-y}, \quad y = 0, 1, \ldots, 10.$$

	B	$\bar{B}$	Total
A	0.10	0.30	0.40
$\bar{A}$	0.40	0.20	0.60
Total	0.50	0.50	1

$$P(Y \geq 3) = p_Y(3) + p_Y(4) + \cdots + p_Y(10)$$

$$= 1 - p_Y(0) - p_Y(1) - p_Y(2)$$

$$= 1 - (0.90)^{10} - \binom{10}{1}(0.10)(0.90)^9 - \binom{10}{2}(0.10)^2(0.90)^8$$

$$\approx 1 - 0.34868 - 0.38742 - 0.19371 = 0.07019.$$

b) If they inspect ***n*** = 10 wells, what is expected value and variance for the number of wells with both impurities? Since $Y \sim \text{Bin}(n = 10, p = 0.10)$,

$$\mu_Y = E[Y] = np = 10(0.10) = 1, \text{ and}$$

$$\sigma_Y^2 = V[Y] = np(1-p) = 0.90.$$

Example B: A gambler will win a game with a probability of 0.45 (and lose otherwise) and all games are independent and identically distributed.

success

a) What is the probability they will win their 3rd game before having their 3rd loss?

e.g. W L W W ✓

L L W W L ✗

$\Rightarrow$ need at least 3 wins in the first 5 games.

Let Y = # of wins in $n = 5$ games $\sim \text{Binomial}(n=5, p=0.45)$.

$$p_Y(y) = \binom{5}{y}(0.45)^y(0.55)^{5-y}, \quad y = 0, 1, \ldots, 5.$$

$$P(Y \geq 3) = p_Y(3) + p_Y(4) + p_Y(5)$$

$$= \binom{5}{3}(0.45)^3(0.55)^2 + \binom{5}{4}(0.45)^4(0.55) + \binom{5}{5}(0.45)^5(0.55)^0$$

$$= 0.27565 + 0.11277 + 0.01845$$

$$= 0.40687.$$

$$p_Y(y) = \binom{5}{y}(0.45)^y(0.55)^{5-y}, \quad y = 0, 1, \ldots, 5.$$

b) Suppose the gambler gets to play the game ***n*** = 5 times and will win an amount, *W*, according to the number of times they win defined below, where *Y* is their number of wins. What is the expected value and variance for *W*?

$$W = \begin{cases} -5+2Y, \text{ if } Y = 0,1,2, \\ \quad 0 \quad\ , \text{ if } Y = 3, \\ 5+3Y\ , \text{ if } Y = 4,5. \end{cases} = \begin{cases} -5, & p_W(-5) = p_Y(0) = 0.55^5 = 0.05033 \\ -3, & p_W(-3) = p_Y(1) = 0.20589 \\ -1, & p_W(-1) = p_Y(2) = 0.33691 \\ 0, & p_W(0) = p_Y(3) = 0.27565 \\ 17, & p_W(17) = p_Y(4) = 0.11277 \\ 20, & p_W(20) = p_Y(5) = 0.01845 \end{cases}$$

$$\mu_W = E[W] = \sum_{\text{all } w} w\, p_W(w) = (-5)(0.05033) + (-3)(0.20589) + \cdots + (20)(0.01845) \approx 1.07986.$$

$$E[W^2] = \sum_{\text{all } w} w^2 p_W(w) = (-5)^2(0.05033) + \cdots + (20)^2(0.01845) = 43.4187.$$

$$\text{Then, } \sigma_W^2 = V[W] = E[W^2] - \mu_W^2 = 43.4187 - 1.07986^2 = 42.2526.$$

c) Suppose $\underbrace{W = 2Y - 4}_{\text{a linear function } Y.}$ What are μ_W and σ_W^2?

$\mu_Y = np = 5(0.45) = 2.25$, $\sigma_Y^2 = np(1-p) = 1.2375$.

- $\mu_W = E[W] = E[2Y - 4] = 2E[Y] - 4 = 2\mu_Y - 4$
  $= 2(2.25) - 4 = 0.50$.

- $\sigma_W^2 = V[W] = V[2Y - 4] = V[2Y] = 4\sigma_Y^2$
  $= 4(1.2375) = 4.95$.

y	w
0	-4
1	-2
2	0
3	2
4	4
5	6

Example C: It is speculated that airlines (sometimes) sell more tickets than available seats as they anticipate some people will not show up. Suppose an airline with ***n*** seats decides to sell ***n*** *+ 2* tickets for a flight.

assume all independent.

a) If they speculate there is a probability of ***p*** that a ticket purchaser will not show up, what is the probability no ticket holder who shows up will be left without a seat?

Let Y = # of $\underbrace{\text{"no shows"}}_{\text{success}}$ from $n+2$ tickets sold.

$$\sim \text{Binomial}(n+2, p), \quad p_Y(y) = \binom{n+2}{y} p^y (1-p)^{n+2-y}, \quad y = 0, 1, \ldots, n+2$$

$$P(Y \geq 2) = 1 - p_Y(0) - p_Y(1) = 1 - (1-p)^{n+2} - (n+2)p(1-p)^{n+1}.$$

OR

Let X = # of people who show up $\sim \text{Bin}(n+2, 1-p)$.

$$p_X(x) = \binom{n+2}{x}(1-p)^x p^{n+2-x}, \quad x = 0, 1, \ldots, n+2.$$

$$P(X \leq n) = 1 - p_X(n+1) - p_X(n+2)$$

$$= 1 - \binom{n+2}{n+1}(1-p)^{n+1} p - \binom{n+2}{n+2}(1-p)^{n+2}$$

$$= 1 - (n+2)(1-p)^{n+1} p - (1-p)^{n+2}.$$

b) Consider one flight with [redacted] 50 seats and the following information

- Tickets are \$500 per seat.
- They estimate there is a probability of 0.03 $\rightarrow p$ that any ticket purchaser will not show up, independent of all others.
- They sell 52 tickets.
- If a ticket holder shows up and the plane is full (meaning 51 or 52 of the ticket holders show up), those ticket holders will be given their money back (the \$500) and additional compensation. $\rightarrow$ call this amount c.

What is the maximum the airline can offer as compensation so that this scenario is profitable (in expectation) for the airline?

$y = 0, \ldots, 52$

Let Y = # of "no shows" $\sim Bin(52, p = 0.03)$, $p_Y(y) = \binom{52}{y}(0.03)^y(0.97)^{52-y}$

$$\text{Let } R = \text{revenue} = \begin{cases} 500(52) = 26000, & \text{if } Y \geq 2 \\ 26000 - 500 - c = 25500 - c, & \text{if } Y = 1 \\ 26000 - 1000 - 2c = 25000 - 2c, & \text{if } Y = 0 \end{cases}$$

$P(R = 26000) = P(Y \geq 2) = 1 - p_Y(0) - p_Y(1) = 0.464845.$

$P(R = 25500 - c) = P(Y = 1) = \binom{52}{1}(0.03)(0.97)^{51} = 0.329977$

$P(R = 25000 - 2c) = P(Y = 0) = (0.97)^{52} = 0.205178$

$$R = \begin{cases} 26000, & \text{prob of } 0.464845 \\ 25500 - c, & \text{" } \quad \text{" } \quad 0.329977 \\ 25000 - 2c, & \text{" } \quad \text{" } \quad 0.205178 \end{cases}$$

$$E[R] = \sum_{\text{all } r} r\,P(R=r) = 26000(0.464845)$$
$$+ (25500 - c)(0.329977) + (25000-2c)(0.205178)$$

$$= 25629.8335 - 0.740333\,c.$$

Note: If they just sell 50 tickets, revenue is $50(500)$
$= 25000$.

$\therefore$ We want $E[R] \geq 25000$.

$$\Rightarrow 25629.8335 - 0.740333c \geq 25000$$

So $c \leq 850.74$.

Example D: Items coming off an assembly line may or may not be defective. Suppose there is a probability of ***p*** that any randomly selected item is defective, independent of any other item. Further, if an item is defective, it may or may not have to be destroyed (some defects can be fixed, and some cannot). Suppose there is a probability of ***q*** that a defective item must be destroyed. What is the probability distribution for the number of items from a random sample of ***n*** items that must be destroyed?

Let X = # of defectives out of n items $\sim Bin(n, p)$.

Consider the # of defectives that must be destroyed

$\Rightarrow Y \mid X = x \sim Bin(x, q)$.

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n.$$

$$P(Y=y \mid X=x) = \binom{x}{y} q^y (1-q)^{x-y}, \quad y = 0, 1, \dots, x.$$

What is $P(Y=y)$? $\Rightarrow$ Partition / Condition over all X.

$$P(Y=y) = P(Y=y \cap X=0) + P(Y=y \cap X=1) + \dots + P(Y=y \cap X=n)$$

$$= \sum_{x=0}^{n} P(Y=y, X=x) = \sum_{x=0}^{n} P(X=x) P(Y=y \mid X=x)$$

$$= \underbrace{\sum_{x=0}^{n}}_{} \left\{ \binom{n}{x} p^x (1-p)^{n-x} \cdot \binom{x}{y} q^y (1-q)^{x-y} \right\}$$

We know $x \geq y$.

Note: $\binom{n}{x}\binom{x}{y} = \frac{n!}{\cancel{x!}\,\cancel{(n-x)!}} \times \frac{\cancel{x!}}{y!\,\cancel{(x-y)!}} \times \binom{n-y}{x-y} \times \frac{\cancel{(x-y)!}\,\cancel{(n-x)!}}{(n-y)!}$

$$= \binom{n-y}{x-y}\binom{n}{y}.$$

$$P(Y=y) = \sum_{x=y}^{n} \binom{n-y}{x-y}\binom{n}{y} p^x (1-p)^{n-x} q^y (1-q)^{x-y}$$

Let $z = x-y$, $x = z+y$

$$= \binom{n}{y} q^y \sum_{z=0}^{n-y} \binom{n-y}{z} p^z p^y (1-p)^{n-y-z} (1-q)^z$$

$$= \binom{n}{y} (pq)^y (1-p)^{n-y} \underbrace{\sum_{z=0}^{n-y} \binom{n-y}{z} \left(\frac{p(1-q)}{1-p}\right)^z}_{}$$

$$= \binom{n}{y}(pq)^y(1-p)^{n-y}\frac{(1-pq)^{n-y}}{(1-p)^{n-y}}$$

$$= \binom{n}{y}(pq)^y(1-pq)^{n-y}, \quad y = 0, 1, \ldots, n.$$

$Y \sim Bin(n, pq)$.

Let $m = n-y$

$$\underbrace{\sum_{z=0}^{m} \binom{m}{z} \left(\frac{p(1-q)}{1-p}\right)^z (1)^{m-z}}_{}$$

Binomial expansion of

$$\left(\frac{p(1-q)}{1-p} + 1\right)^m = \left(\frac{1-pq}{1-p}\right)^{n-y}$$

<u>Example E – The Matching Game:</u> A room full of ***n*** students all throw their *OneCards* in a hat. Then, they each draw one *OneCard* at random from the hat and hold on to it. If a student selects their own card it is considered a match.

a) What is the probability distribution for the number of matches? $n = 2, 3, 4$

b) What is the expected number of matches? $\Rightarrow$ 1 for any n.

c) What is the variance for the number of matches? $\Rightarrow$ 1 for any n.

Let the random variable Y denote the number of matches. Consider approximations for the questions above using the Binomial distribution.

<u>Recall</u>: Let $X_i := \begin{cases} 1, & \text{if } i\text{th person selects own card} \\ 0, & \text{otherwise} \end{cases}, \quad i = 1, \dots, n.$

$\sim \text{Bernoulli}(p = 1/n)$.

Then Y = total # of matches $= X_1 + X_2 + \cdots + X_n$.

<u>IF</u> $X_1, \dots, X_n$ are independent then $Y \sim \text{Bin}(n, p = 1/n)$.

$\Rightarrow$ However they are not independent.

e.g. $P(X_2 = 1 \mid X_1 = 1) = \frac{1}{n-1} \neq P(X_2 = 1) = \frac{1}{n}$

But if n is large,

$\frac{1}{n-1} \approx \frac{1}{n}$.

approx.

So for large n, $Y \approx \text{Bin}(n, p = 1/n)$.

- $\mu_Y = E[Y] = np = n(1/n) = 1$ (the exact value)
- $\sigma_Y^2 = V[Y] = np(1-p) = \frac{n-1}{n}$ ($\to 1$ as $n \to \infty$)

$$P(Y = y) \approx \binom{n}{y}\left(\frac{1}{n}\right)^y\left(\frac{n-1}{n}\right)^{n-y}, \quad y = 0, 1, 2, \ldots, n-1, n$$

e.g. $n = 4$:

Actual: $p_Y(0) = 9/24 \approx 0.375$, $p_Y(1) = 8/24 \approx 0.333$,

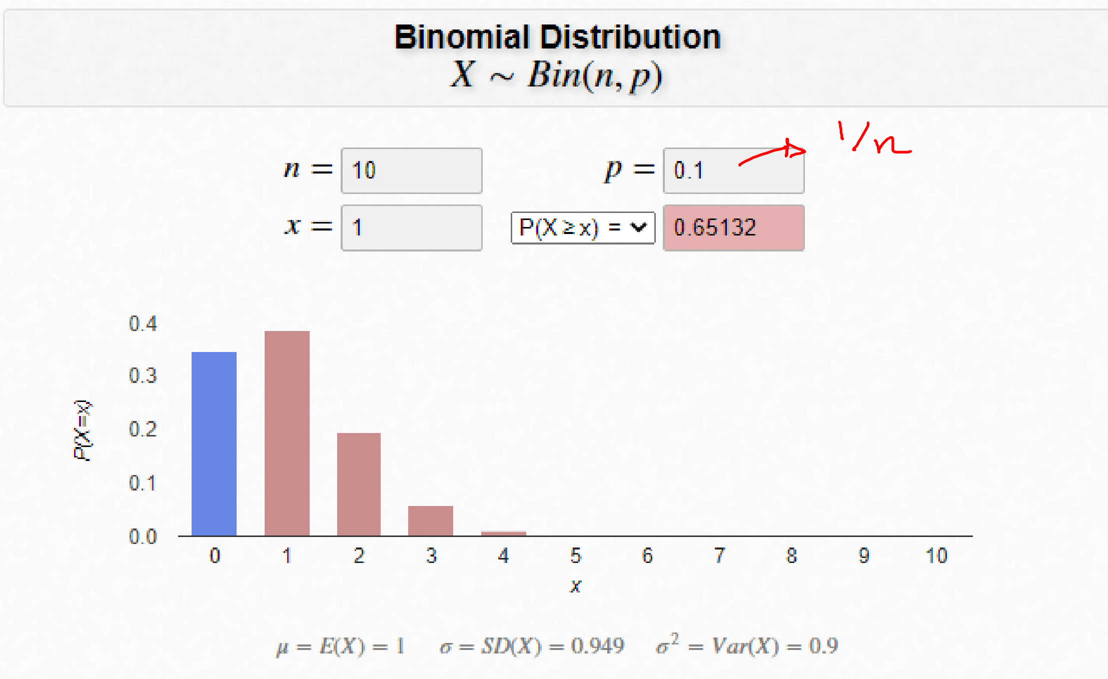
$p_Y(2) = 6/24 \approx 0.25$, $p_Y(4) = 1/24 \approx 0.04167$.

Bin. Approx: $Y \approx \text{Bin}(n=4, p = 1/4)$, $p_Y(y) = \binom{4}{y}\left(\frac{1}{4}\right)^y\left(\frac{3}{4}\right)^{4-y}$, $y = 0, 1, \ldots, 4$.

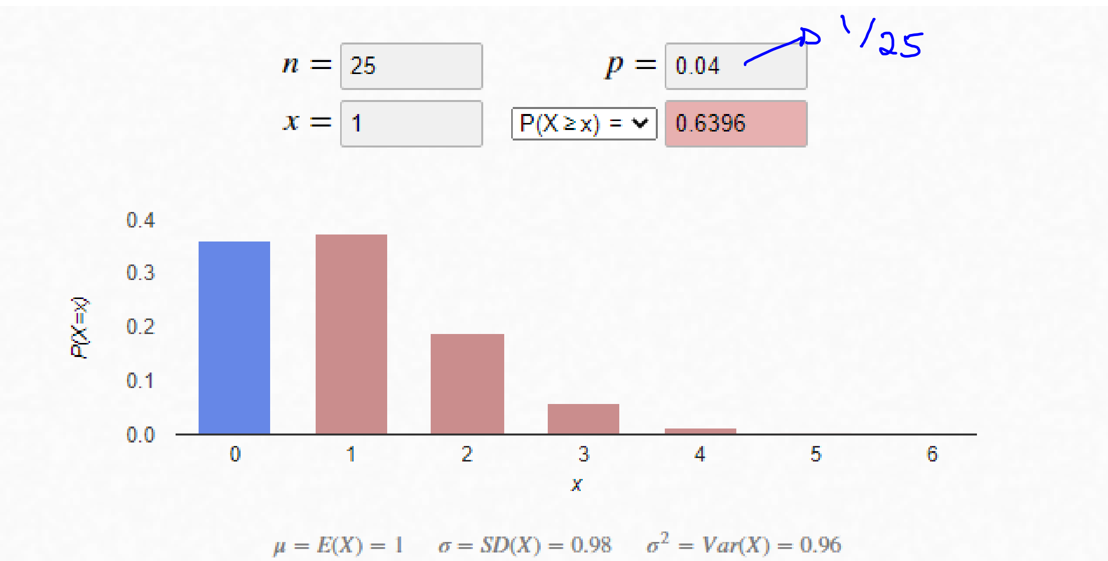
$p_Y(0) \approx 0.3164$, $p_Y(1) \approx 0.4219$, $p_Y(2) \approx 0.2109$,

$p_Y(3) \approx 0.04688$, $p_Y(4) \approx 0.003906$.

(Not that great .... but n is only 4.)



$1/n$

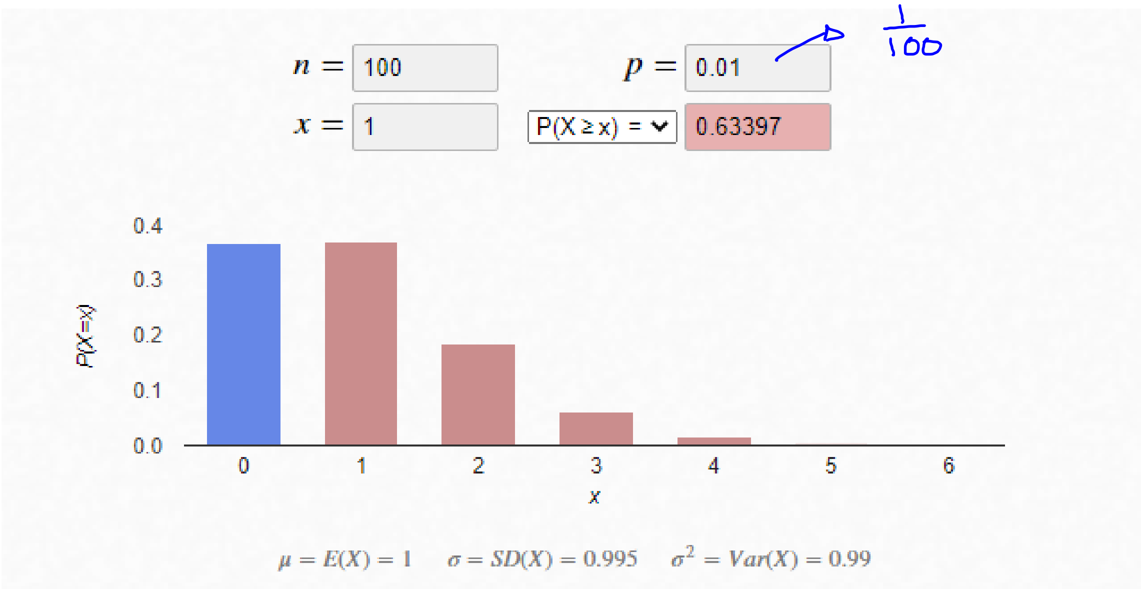


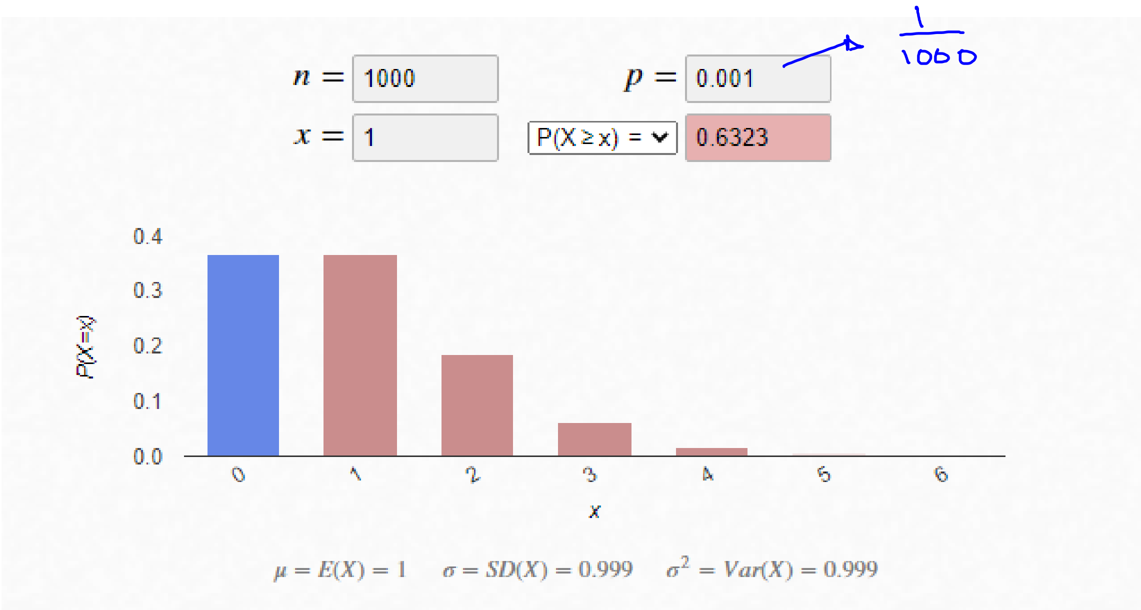
$1/25$

Recall: $P(Y \geq 1)$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots$$

$$\approx 1 - \frac{1}{e} \approx 0.63212.$$





https://homepage.divms.uiowa.edu/~mbognar/applets/bin.html