University of Waterloo W. H. Cherry

DISCRETE PROBABILITY DISTRIBUTION HISTOGRAMS: Hypergeometric, Binomial, Negative binomial, Poisson

This Highlight #4 is concerned with visualizing probability, a topic pursued in Statistical Highlight #5. Here, the emphasis is on histograms for four discrete distributions whuich are often introduced in introductory probability courses. Statistical issues arising from probability distributions are taken up in Statistical Highlight #2 on the lower half of page HL2.18.

Distn. 1. The **hypergeometric** distribution can be used to model the process of selecting a sample of n balls equiprobably with*out* replacement from an urn containing a total of $\mathbf{N}$ balls of which $\mathbf{R}$ are red and $\mathbf{N}-\mathbf{R}$ are black. If the random variable R represents the number of red balls in the sample, its probability function is as given in equation (HL4.1) at the right ('urn' here corresponds to 'population' in statistics).

$$f(r) = \frac{\binom{\mathbf{R}}{r}\binom{\mathbf{N}-\mathbf{R}}{\mathrm{n}-r}}{\binom{\mathbf{N}}{\mathrm{n}}}, \quad r = 0, 1, 2, ..., \mathrm{n}. \qquad \text{-----(HL4.1)}$$

When $\mathbf{N}=10$, $\mathbf{R}=4$ or $\mathbf{R}=6$, and $\mathrm{n}=6$, we can tabulate and graph $f(r)$ as follows:

Two Hypergeometric Probability Functions

Table HL4.1

r	$f(r)$
0	$\frac{1}{210} \simeq 0.0048$
1	$\frac{24}{210} \simeq 0.1143$
2	$\frac{90}{210} \simeq 0.4286$
3	$\frac{80}{210} \simeq 0.3810$
4	$\frac{15}{210} \simeq 0.0714$
5	0
6	0

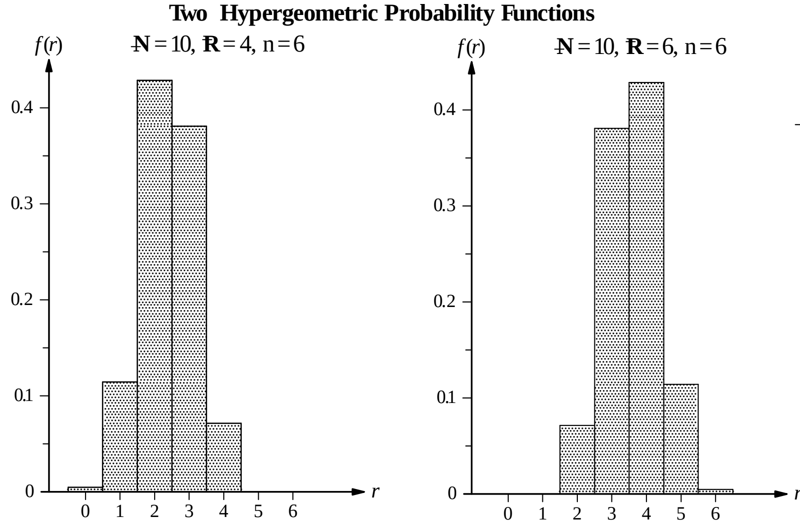


Table HL4.2

r	$f(r)$
0	0
1	0
2	$\frac{15}{210} \simeq 0.0714$
3	$\frac{80}{210} \simeq 0.3810$
4	$\frac{90}{210} \simeq 0.4286$
5	$\frac{24}{210} \simeq 0.1143$
6	$\frac{1}{210} \simeq 0.0048$

Distn. 2. The **binomial** distribution can be used to model a process called a Bernoulli trial, which has the following (restricted) characteristics:

- a Bernoulli trial has *two* outcomes (generically called 'success' and 'failure');
- under repetition, the outcomes of Bernoulli trials ('success' and 'failure') are *probabilistically independent* events;
- under repetition, the probability of 'success' (and, hence, the probability of 'failure') is *constant* [π say, and $1-\pi$].

An example of Bernoulli trials is selecting equiprobably from an urn of red and black balls *with* replacement [to be contrasted with the hypergeometric distribution for selecting with*out* replacement in Dist. 1 above]; a more familiar example of a Benoulli trial is tossing a coin.

The binomal probability function (for the random variable Y) is:

$$\Pr(y \text{ 'successes' in n trials}) = \binom{\mathrm{n}}{y}\pi^{y}(1-\pi)^{\mathrm{n}-y}; \quad y = 0, 1, 2, ..., \mathrm{n}. \qquad \text{-----(HL4.2)}$$

When $\mathrm{n}=10$ and π successively takes the values 0.2, 0.5 and 0.9, the probability functions [$f(x)$, $f(y)$, $f(z)$] of the respective random variables X, Y and Z are tabulated and graphed as shown on page HL4.3.

Distn. 3. The **negative binomial** distribution is also used to model repetition of Bernoulli trials, but it differs from the binomial distribution in its concern with the probability that the r*th* 'success' occurs on the n*th* trial or, equivalently, that v 'failures' occur before the r*th* 'success'. The negative binomial probability function (for random variable V) is:

$$\Pr(v \text{ 'failures' before the r}th\text{ 'successes'}) = \binom{\mathrm{r}+v-1}{\mathrm{r}-1}\pi^{\mathrm{r}}(1-\pi)^{v} \quad ; \quad v = 0, 1, 2, 3, ... \qquad \text{-----(HL4.3)}$$

When $\mathrm{r}=2$ or $\mathrm{r}=5$ and $\pi=0.4$, the probability functions $f(v)$ and $f(w)$ are tabulated and graphed overleaf on the lower half of page HL4.2. [Terms of these infinite sets of values are evaluated to $V=15$ and $W=25$ to show the behaviour of the right tail of these distributions.]

Distn. 4. The **Poisson** distribution is used to model processes involving 'outcomes occurring unpredictably (on an *individual* basis) in space or time'; the more common shorter description of modelling *random events in space or time* has the *dis*advantage of using 'random' and 'event' in their colloquial (*not* their probabilistic) sense.

The derivation of the Poisson probability function involves three assumptions:

- *independence*: events (which are *numbers of outcomes in an interval*) are *probabilistically independent* for *non-*

(*continued overleaf*)

Distn. 4. (cont.)

- *independence*: overlapping (spatial or temporal) intervals; the shorter version of this assumption – the *numbers* of outcomes comprising events in non-overlapping intervals are independent – risks confusing probabilistic independence of events with independence of numbers.
- *individuality*: outcomes occur singly, not in groups; *i.e.*, the probability of more than one outcome in a sufficiently small interval is negligible;
- *homogeneity*: outcomes occur at a uniform rate over the entire interval under consideration so that, in a (sub-)interval of length t, the expected number of outcomes is λt, where the *intensity parameter* λ is the *avrage rate of outcomes over the entire interval.*

  This *mathematical* assumption has troublesome real-world implications – a *uniform* underlying process exhibiting *unpredictable* actual individual outcomes.

The Poisson probability function (for the random variable S) is:

$$\Pr(s \text{ outcomes in an interval of length t}) = \frac{(\lambda \mathrm{t})^s e^{-\lambda \mathrm{t}}}{s!} = \frac{\mu^s e^{-\mu}}{s!} \qquad s = 0, 1, 2, 3, \ldots.. \qquad \text{-----(HL4.4)}$$

In contrast to Bernoulli trials where *a person* is usually involved in repetition (*e.g.*, in coin tossing), a Poisson process often repeats with*out* human intervention and the 'repetitions' arise from the (spatial or temporal) intervals *imposed* on the outcomes.

When μ takes values of 1, 5 and 10, the probability functions [$f(s)$, $f(t)$ and $f(u)$] of the three random variables S, T and U are tabulated and graphed as shown on the page HL4.4. [Terms of these infinite sets of values are evaluated out to $S = 10$, $T = 15$ and $U = 20$, respectively, to show the behaviour of the right tails of the distributions.]

Distn. 3. (cont.) Two negative binomial distribution probability functions (r = 2, π = 0.4 and (r = 5, π = 0.4):

Table HL4.3

v	$f(v)$
0	0.16
1	0.192
2	0.172 8
3	0.138 24
4	0.103 68
5	0.074 649 6
6	0.052 254 72
7	0.035 831 808
8	0.024 186 470
9	0.016 124 314
10	0.010 642 047
11	0.006 965 703
12	0.004 527 707
13	0.002 925 595
14	0.001 880 740
15	0.001 203 674
	(0.997 912 378)

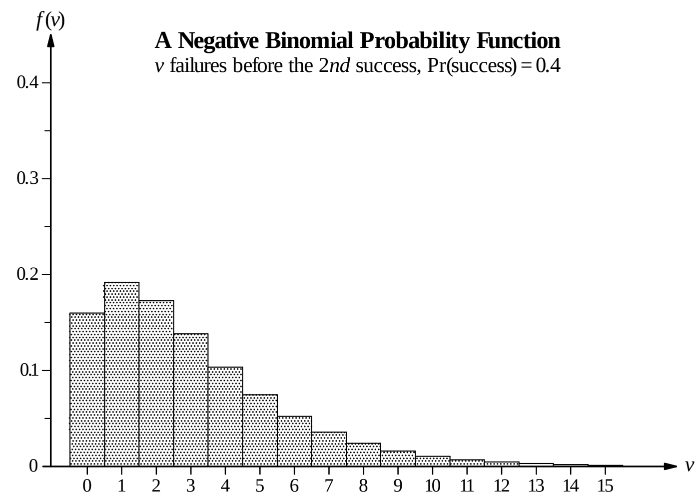
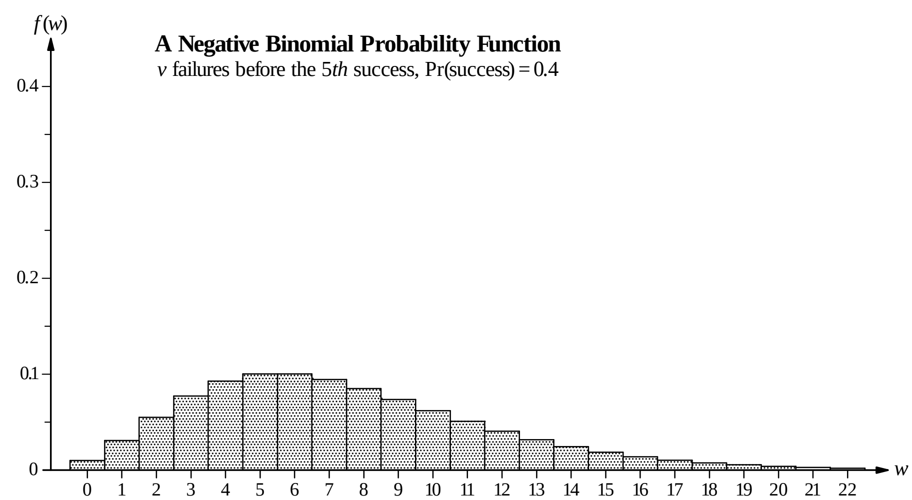


Table HL4.4

v	$f(w)$	w	$f(w)$
0	0.010 24	16	0.013 996 316
1	0.030 72	17	0.010 373 740
2	0.055 296	18	0.007 607 409
3	0.077 414 4	19	0.005 525 382
4	0.092 897 28	20	0.003 978 275
5	0.100 329 062	21	0.002 841 625
6	0.100 329 062	22	0.002 014 970
7	0.094 595 973	23	0.001 419 240
8	0.085 136 325	24	0.000 993 468
9	0.073 784 859	25	0.001 691 454
10	0.061 979 281		(0.998 489 925)
11	0.050 710 321		
12	0.040 568 257		
13	0.031 830 478		
14	0.024 554 940		
15	0.018 661 754		
.	(0.949 048 046)		



(*continued*)

University of Waterloo W. H. Cherry

DISCRETE PROBABILITY DIST...N HISTOGRAMS: Hypergeometric, Binomial, Negative binomial, Poisson (continued 1)

Distn. 2. (cont.) Three binomial distribution probability functions (n = 10, $\pi = 0.2$, $\pi = 0.5$ and $\pi = 0.9$):

Table HL4.5

w	$f(w)$
0	0.107 374 182
1	0.268 435 456
2	0.301 989 888
3	0.201 326 592
4	0.088 080 384
5	0.026 424 115
6	0.005 505 024
7	0.000 786 432
8	0.000 073 728
9	0.000 004 096
10	0.000 000 102
	(1.)

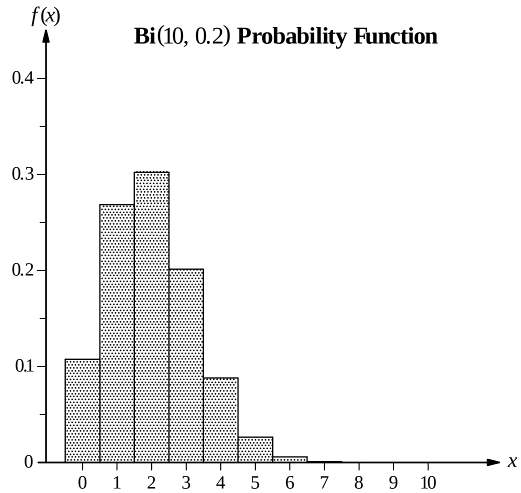


Table HL4.6

y	$f(y)$
0	0.000 976 563
1	0.009 765 625
2	0.043 945 313
3	0.117 187 500
4	0.205 078 125
5	0.246 093 750
6	0.205 078 125
7	0.117 187 500
8	0.043 945 313
9	0.009 765 625
10	0.000 976 563
	(1.)

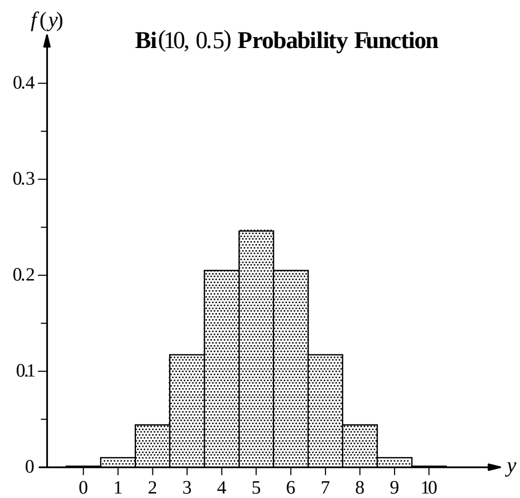
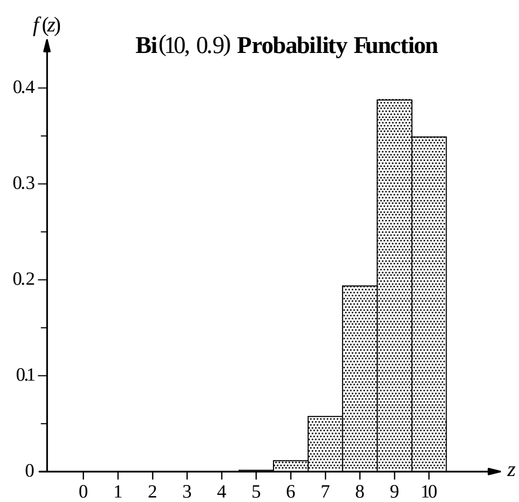


Table HL4.7

z	$f(z)$
0	0.000 000 000
1	0.000 000 009
2	0.000 000 365
3	0.000 008 748
4	0.000 137 781
5	0.001 488 035
6	0.011 160 261
7	0.057 395 628
8	0.193 710 245
9	0.387 420 489
10	0.348 678 440
	(1.)



NOTE: 1. We see that the most probable value of each random variable, corresponding to the *highest* bar of each histogram, is given by the product, n $\times \pi$, of the binomial distribution parameters; this is (of course) the *mean* of the distribution.

 (*continued overleaf*)

Distn. 4. Three Poisson distribution probability functions ($\lambda t \equiv \mu = 1$, $\lambda t \equiv \mu = 5$ and $\lambda t \equiv \mu = 10$):
(cont.)

Table HL4.8

s	$f(s)$
0	0.367 879 441
1	0.367 879 441
2	0.183 939 721
3	0.061 313 240
4	0.015 328 310
5	0.003 065 662
6	0.000 510 944
7	0.000 072 992
8	0.000 009 124
9	0.000 001 014
10	0.000 000 101
	(0.999 999 990)

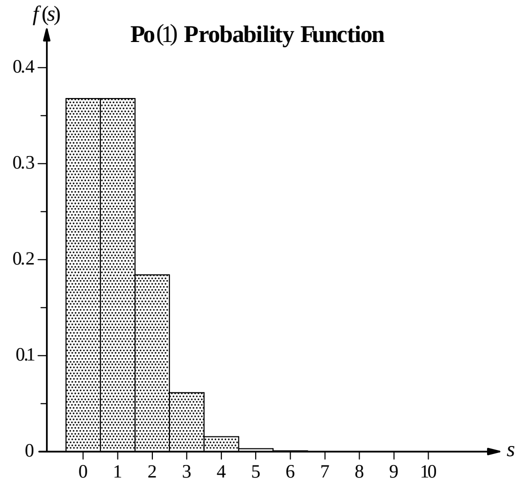


Table HL4.9

s	$f(t)$	s	$f(t)$
0	0.006 737 947	11	0.008 242 177
1	0.033 689 735	12	0.003 434 240
2	0.084 224 337	13	0.001 320 862
3	0.140 373 895	14	0.000 471 736
4	0.175 467 369	15	0.000 157 245
5	0.175 067 369		(0.999 930 991)
6	0.146 222 808		
7	0.104 444 863		
8	0.065 278 039		
9	0.036 265 577		
10	0.018 132 788		
	(0.986 304 731)		

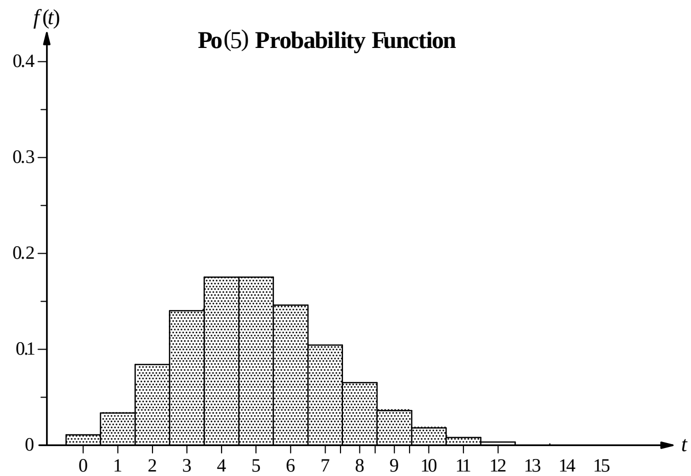
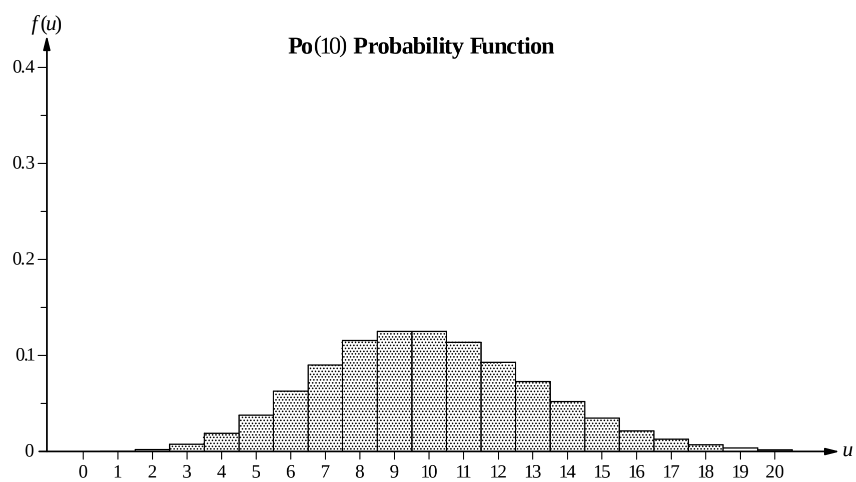


Table HL4.10

s	$f(u)$	s	$f(u)$
0	0.000 045 400	11	0.113 736 396
1	0.000 453 999	12	0.094 780 330
2	0.002 269 996	13	0.072 907 946
3	0.007 566 655	14	0.052 077 104
4	0.018 916 637	15	0.034 718 070
5	0.037 833 275	16	0.021 698 794
6	0.063 055 458	17	0.012 763 996
7	0.090 079 226	18	0.007 091 109
8	0.112 599 032	19	0.003 732 163
9	0.125 110 036	20	0.001 866 081
10	0.125 110 036		
	(0.998 411 739)		(0.986 304 731)



NOTES: 2. We see that the most probable value of the random variable, corresponding to the *highest* bar of each histogram, is the integer part of the value of the parameter (here, the mean μ); when this value is an integer, the bar to the *left* of this tallest bar is also of the same height (*i.e.*, there are *two* most probable values when μ is an integer).

3. Th asymmetry of the Po(1) distribtiion is substantially decreased in the Po(5) probability function and it continues to decrease as μ increases. The relatively symmetrical Po(10) distribution is approaching *'bell-shaped,'* reminding us of the *normal* approximation for finding values for (sums of) Poisson probabilities when μ is large (say 20 or more).