ST 430
Homework #10 Solutions

10.2.2 Let X_1 = number of round and yellow phenotypes, X_2 = number of round and green phenotypes, and so on. Then $P(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1) =$
$\frac{4!}{1!1!1!1!}\left(\frac{9}{16}\right)^1\left(\frac{3}{16}\right)^1\left(\frac{3}{16}\right)^1\left(\frac{1}{16}\right)^1 = 0.0297.$

10.2.4 Let Y denote a recruit's IQ and let X_i denote the number of recruits in class i, $i = 1, 2, 3$. Then $p_1 = P(\text{class I}) = P(Y < 90) = P\left(Z < \frac{90-100}{16}\right) = 0.2643$, $p_2 = P(\text{class II}) = P(90 \le Y \le 110) =$ $P\left(\frac{90-100}{16} \le Z \le \frac{110-100}{16}\right) = 0.4714$, and $p_3 = P(\text{class III}) = P(Y > 110) = 1 - p_1 - p_2 =$ 0.2643. From Theorem 10.2.1, $P(X_1 = 2, X_2 = 4, X_3 = 1) = \frac{7!}{2!4!1!}(0.2643)^2(0.4714)^2(0.2643)^1 =$ 0.0957.

10.2.8 $M_{X_1,X_2,X_3}(t_1,t_2,t_3) = \sum\sum\sum e^{t_1k_1+t_2k_2+t_3k_3} \cdot \frac{n!}{k_1!k_2!k_3!} \cdot p_1^{k_1}p_2^{k_2}p_3^{k_3} =$
$\sum\sum\sum \frac{n!}{k_1!k_2!k_3!}\left(p_1e^{t_1}\right)^{k_1}\left(p_2e^{t_2}\right)^{k_2}\left(p_3e^{t_3}\right)^{k_3}$, where the summation extends over all the values of (k_1, k_2, k_3) such that $k_i \ge 0$, $i = 1, 2, 3$ and $k_1 + k_2 + k_3 = n$. Recall Newton's binomial expansion. Applied here, it follows that the triple sum defining the moment-generating function for (X_1, X_2, X_3) can also be written $\left(p_1e^{t_1} + p_2e^{t_2} + p_3e^{t_3}\right)^n$.

10.2.10 The log of the likelihood vector $(k_1, k_2, \ldots, k_t)$ is $\log L = \log\ p_1^{k_1}p_2^{k_2}\ldots p_t^{k_t} = k_1 \log p_1 +$ $k_2 \log p_2 + \ldots + k_t \log p_t$, where the p_i's are constrained by the condition that $\sum_{i=1}^{t} p_i = 1$. Finding the MLE for the p_i's can be accomplished using Lagrange multipliers. Differentiating $\log L - \lambda\sum_{i=1}^{t} p_i$ with respect to each p_i gives $\frac{\partial}{\partial p_i}\left[\log L - \lambda\sum_{i=1}^{t} p_i\right] = \frac{k_i}{p_i} - \lambda$, $i = 1, 2, \ldots, t$. But these derivatives equal 0 only if $\frac{k_i}{p_i} = \lambda$ for all i. The latter equations, together with the fact that $\sum_{i=1}^{t} p_i = 1$, imply that $\hat{p}_i = \frac{k_i}{n}$, $i = 1, 2, \ldots, t$.

10.3.2 If the hypergeometric model applies, $\pi_1 = P$(0 whites are drawn) = $\binom{4}{0}\binom{6}{2}\Big/\binom{10}{2} = \frac{15}{45}$, π_2 = P(1 white is drawn) = $\binom{4}{1}\binom{6}{1}\Big/\binom{10}{2} = \frac{24}{45}$, and $\pi_3 = P$(2 whites are drawn) = $\binom{4}{2}\binom{6}{0}\Big/\binom{10}{2} = \frac{6}{45}$. Let p_1, p_2, and p_3 denote the actual probabilities of drawing 0, 1, and 2 white chips, respectively. To test H_0: $p_1 = \frac{15}{45}$, $p_2 = \frac{24}{45}$, $p_3 = \frac{6}{45}$ versus H_1: at least one $p_i \neq \pi_i$, reject H_0 if $d \geq \chi^2_{1-\alpha,k-1} = \chi^2_{.90,2} = 4.605$.

Here, $d = \frac{(35 - 100(15/45))^2}{100(15/45)} + \frac{(55 - 100(24/45))^2}{100(24/45)} + \frac{(10 - 100(6/45))^2}{100(6/45)} = 0.96$, so H_0 (and the hypergeometric model) would not be rejected.

10.3.4 If births occur randomly in time, then $\pi_1 = P$(baby is born between midnight and 4 A.M.) = $\frac{1}{6}$ and $\pi_2 = P$(baby is born at a "convenient" time) = $1 - \pi_1 = \frac{5}{6}$. Let p_1 and p_2 denote the actual probabilities of birth during those two time periods. The null hypothesis to be tested is H_0: $p_1 = \frac{1}{6}$, $p_2 = \frac{5}{6}$. At the $\alpha = 0.05$ level of significance, H_0 should be rejected if $d \geq \chi^2_{.95,1}$ = 3.841. Given that $n = 2650$ and that X_1 = number of births between midnight and 4 A.M. = 494, it follows that $d = \frac{(494 - 2650(1/6))^2}{2650(1/6)} + \frac{(2156 - 2650(5/6))^2}{2650(5/6)} = 7.44$. Since the latter exceeds 3.841, we reject the hypothesis that births occur uniformly in all time periods.

10.3.6 In the terminology of Theorem 10.3.1, $X_1 = 1383$ = number of schizophrenics born in first quarter and X_2 = number of schizophrenics born after the first quarter. By assumption, $n\pi_1$ = 1292.1 and $n\pi_2 = 3846.9$ (where $n = 5139$). The null hypothesis that birth month is unrelated to schizophrenia is rejected if $d \geq \chi^2_{.95,1} = 3.841$. But $d = \frac{(1383 - 1292.1)^2}{1292.1} + \frac{(3756 - 3846.9)^2}{3846.9} = 8.54$, so H_0 is rejected, suggesting that month of birth may, indeed, be a factor in the incidence of schizophrenia.

10.3.10 Let $p_i = P$(horse starting in post position i wins), $i = 1, 2, \ldots, 8$. One relevant null hypothesis to test would be that p_i is not a function of i—that is, H_0: $p_1 = p_2 = \ldots = p_8 = \frac{1}{8}$ versus H_1: at least one $p_i \neq \frac{1}{8}$. If $\alpha = 0.05$, H_0 should be rejected if $d \geq \chi^2_{.95,7} = 14.067$. Each $E(X_i)$ in this case is $144 \cdot \frac{1}{8} = 18.0$, so $d = \frac{(32 - 18.0)^2}{18.0} + \frac{(21 - 18.0)^2}{18.0} + \ldots + \frac{(11 - 18.0)^2}{18.0} = 18.72$. Since $18.72 \geq \chi^2_{.95,7}$, we reject H_0 (which is not surprising because faster horses are typically awarded starting positions close to the rail).

10.3.12 Let the random variable Y denote the prison time served by someone convicted of grand theft auto. In the accompanying table is the frequency distribution for a sample of 50 y_i's, together with expected frequencies based on the null hypothesis that $f_Y(y) = \frac{1}{9}y^2$, $0 \le y \le 3$. For example, $E(X_1) = 50 \cdot \pi_1 = 50\int_0^1 \frac{1}{9}y^2 dy = 1.85$. Combining the first two intervals (because $E(X_1) < 5$) yields $k = 2$ final classes, so H_0: $f_Y(y) = \frac{1}{9}y^2$, $0 \le y \le 3$ should be rejected if $d \ge \chi^2_{.95,1} = 3.841$. But $d = \frac{(24-14.81)^2}{14.81} + \frac{(26-35.19)^2}{35.19} = 8.10$, implying that the proposed quadratic pdf does not provide a good model for describing prison time.

Prison time, y	Freq.	π_i	$E(X_i)$	
$0 \le y < 1$	8	1/27	1.85	} 14.81
$1 \le y < 2$	16	7/27	12.96	
$2 \le y < 3$	26	19/27	35.19	
	50	1	50.00	

10.4.2 For the Poisson pdf, $\hat{\lambda} = \frac{59(0)+27(1)+9(2)+1(3)}{96} = 0.50$ so the hypotheses being tested are H_0: $P(i \text{ vacancies}) = e^{-0.50}(0.50)^i/i!$, $i = 0, 1, 2, \ldots$ vs. H_1: $P(i \text{ vacancies}) \ne e^{-0.50}(0.50)^i/i!$, $i = 0, 1, 2, \ldots$ As the table indicates, the original frequency distribution needs to have several classes combined because the expected frequencies are too small.

No. of vacancies, i	No. of years	$\hat{p}_i$	$96 \cdot \hat{p}_i$	
0	59	0.607	58.27	
1	27	0.303	29.09	
2	9	0.076	7.30	
3	1	0.013	1.25	} 8.65
4+	0	0.001	0.10	
	96	1.000	96.00	

10.4.4 Let $\hat{\lambda} = \frac{109(0)+65(1)+22(2)+3(3)+4(1)}{200} = 0.61$. Then the model to be fit under H_0 is the Poisson pdf, $p_X(i) = e^{-0.61}(0.61)^i/i!$, $i = 0, 1, 2, \ldots$ Using $t = 4$ final classes (the combined "4.8" is close enough to 5 for the χ^2 approximation to be adequate), we should reject H_0 if $d_1 \ge \chi^2_{.99,4-1-1} = 9.210$. In the table, the observed and expected frequencies are in excellent agreement, so d_1 will be very small (and the Poisson model will not be rejected). Specifically, $d_1 = \frac{(109-108.7)^2}{108.7} + \frac{(65-66.3)^2}{66.3} + \frac{(22-20.2)^2}{20.2} + \frac{(4-4.8)^2}{4.8} = 0.32$.

No. of Deaths, i	Freq.	$\hat{p}_i$	$200 \cdot \hat{p}_i$	
0	109	0.5434	108.7	
1	65	0.3314	66.3	
2	22	0.1011	20.2	
3	3	0.0206	4.1	} 4.8
4+	1	0.0035	0.7	
	200	1.0000	200.0	

10.3.9 Let the random variable X denote the length of a World Series. Then $P(X = 4) = \pi_1 = P$(AL wins in 4) + P(NL wins in 4) = 2 · P(AL wins in 4) = $2\left(\frac{1}{2}\right)^4 = \frac{1}{8}$. Similarly, $P(X = 5) = \pi_2$ = 2 · P(AL wins in 5) = 2 · P(AL wins exactly 3 of first 4 games) · P(AL wins 5th game) = $2 \cdot \binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1 \cdot \frac{1}{2} = \frac{1}{4}$. Also, $P(X = 6) = \pi_3$ = 2· P(AL wins exactly 3 of first 5 games) · P(AL wins 6th game) = $2 \cdot \binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right) = \frac{5}{16}$, and $P(X = 7) = \pi_4 = 1 - P(X = 4) - P(X = 5) - P(X = 6) = \frac{5}{16}$. Listed in the table is the information necessary for calculating the goodness-of-fit statistic d. The "Bernoulli model" is rejected if $d \geq \chi^2_{.90,3} = 6.251$. For these data, $d = \frac{(9-6.25)^2}{6.25} + \frac{(11-12.50)^2}{12.50} + \frac{(8-15.625)^2}{15.625} + \frac{(22-15.625)^2}{15.625} = 7.71$, so H_0 is rejected.

Number of games	Number of years	$50 \cdot \pi_i$
4	9	6.25
5	11	12.50
6	8	15.625
7	22	15.625
	50	50.000