Statistics (II) Exam II, June 6, 2002

(1) (a) A student got 9207 heads and 8743 tails in 17,950 coin tosses. Is this a significant discrepancy from the null hypothesis $H_0: \ p = \frac{1}{2}$?
(b) If the student performed the experiment by tossing groups of five coins at a time, with the following results.

Are the data consistent with the hypothesis that all the coins were fair ($p = \frac{1}{2}$) ?
(c) Are the data consistent with the hypothesis that all five coins had the same probability of heads but that this probability was not necessary $\frac{1}{2}$?

(2) Show that the transformation $Y = \sin^{-1}\sqrt{\hat{p}}$ is variance-stabilizing if $\hat{p} = \frac{X}{n}$, where $X \sim bin(n, p)$.

(3) The Cauchy distribution has the probability density function $f(x) = \frac{1}{\pi}(\frac{1}{1+x^2})$, $-\infty < x < \infty$ and the double exponential distribution has the probability density function $g(x) = \frac{1}{2}e^{-|x|}, -\infty < x < \infty$. What would be the qualitative shapes of normal probability plots of samples from Cauchy and double exponential distributions, respectively.

(4) Let n_j denote the observed count and $\hat{n}_j$ denote the expected count of a hanging histogram. If we apply a variance-stabilizing transformation $f(x) = \sqrt{x}$, then $E(\sqrt{n_j}) \cong \sqrt{np_j}$, $Var(\sqrt{n_j}) \cong \frac{1}{4}$. Explain how to assess the significance of the deviation $\sqrt{n_j} - \sqrt{\hat{n}_j}$.

(5) What proportion of the observations from a normal sample would you expect to be marked by an asterisk on a boxplot.

(6) For a sample of size $n = 3$ from a continuous probability distribution. what is $P(X_{(1)} < \eta < X_{(3)})$, where η is the median of the distribution.

(7) Consider the two c.d.f.'s $F(x) = x, 0 \le x \le 1, G(x) = x^2, 0 \le x \le 1$. Sketch a Q-Q plot of F versus G.

(8) Let $X_1, X_2, \cdots, X_n$ be a sample (i.i.d.) from a distribution function F and let F_n denote the empirical cdf. Show that $Cov(F_n(u), F_n(v)) = \frac{1}{n}[F(m) - F(u)F(v)]$, where $m = \min(u, v)$. Are $F_n(u)$ and $F_n(v)$ positively correlated or negatively correlated ?

(9) States four measures of location and four measures of dispersion.