STAT 341 ; Final Exam: Useful facts. Mar 17, 2010

1. Permutations and combinations

There are $n! = \prod_{i=1}^{n} i = 1.2.3.4.\ldots n$ permutations of n objects.

There are $\binom{n}{k} = n!/(k!(n-k)!)$ ways of choosing a given k objects from n.

2. Joint and conditional probabilities

If C and D are any events: $\mathrm{P}(C \bigcup D) = \mathrm{P}(C) + \mathrm{P}(D) - \mathrm{P}(C \bigcap D)$.

The conditional probability of C given D is $\mathrm{P}(C \mid D) = \mathrm{P}(C \bigcap D) / \mathrm{P}(D)$.

C and D are independent if $\mathrm{P}(C \bigcap D) = \mathrm{P}(C).\mathrm{P}(D)$.

3. Laws and theorems

Suppose $E_1, \ldots, E_k$ is a partition of Ω. That is $E_i \bigcap E_j$ is empty, and $E_1 \bigcup E_2 \bigcup \ldots \bigcup E_k = \Omega$.

The law of total probability states that: $\mathrm{P}(D) = \sum_{j=1}^{k} \mathrm{P}(D \bigcap E_j) = \sum_{j=1}^{k} \mathrm{P}(D \mid E_j)\, \mathrm{P}(E_j)$

Bayes' Theorem states that: $\mathrm{P}(E_i \mid D) = \mathrm{P}(D \mid E_i)\, \mathrm{P}(E_i)/\mathrm{P}(D)$

4. Random variables and distributions

	discrete (mass)	continuous (density)
Probability mass/density function	pmf: $\mathrm{P}(X = x) = p_X(x)$	pdf: $f_X(x)$
Cumulative dist. func. CDF, $\mathrm{P}(X \leq x)$	$F_X(x) = \sum_{w \leq x} p_X(w)$	$F_X(x) = \int_{-\infty}^{x} f_X(w) dw$
Joint mass/density func. of (X, Y)	$p_{X,Y}(x, y) = \mathrm{P}(X = x, Y = y)$	$f_{X,Y}(x, y)$
Marginal mass/density of X	$p_X(x) = \sum_y p_{X,Y}(x, y)$	$f_X(x) = \int_{y=-\infty}^{\infty} f_{X,Y}(x, y) dy$
Conditional of X given $Y = y$	$p_X(x \mid Y = y) = p_{X,Y}(x, y)/p_Y(y)$	$f_X(x \mid Y = y) = f_{X,Y}(x, y)/f_Y(y)$
Independence of X and Y	$p_{X,Y}(x, y) = \mathrm{P}(X = x)\mathrm{P}(Y = y)$	$f_{X,Y}(x, y) = f_X(x) f_Y(y)$

5. Moments of random variables:

Expectation: $\mathrm{E}(g(X)) = \sum_x g(x)\ \mathrm{P}(X = x)$ $\qquad \int_{-\infty}^{\infty} g(x) f_X(x) dx$

provided the sum/integral converges absolutely.

For any random variables X and Y:

Variance: $\mathrm{var}(X) = \mathrm{E}((X - E(X))^2) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$

Always: $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$, $\quad \mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$, $\quad \mathrm{var}(aX + b) = a^2\ \mathrm{var}(X)$.

If X and Y are independent: $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y)$

If $X_1, ..., X_n$ are i.i.d, $\overline{X_n} \equiv (1/n) \sum_{i=1}^{n} X_i$ then $\mathrm{E}(\overline{X_n}) = \mathrm{E}(X_1)$, $\mathrm{var}(\overline{X_n}) = \mathrm{var}(X_1)/n$.

6. Basics of estimation of parameter θ

(a) An n-sample is a set of n independent data random variables, $Y_1, ..., Y_n$ all from the same distribution (i.i.d), indexed by unknown parameter(s) θ.

(b) A *statistic*, T, is any function of $Y_1, ..., Y_n$.

(c) An *estimator* of $g(\theta)$ is a statistic used to estimate $g(\theta)$.

(d) The *estimate* of $g(\theta)$ is the value taken by the estimator in any particular instance.

(e) The first Method of Moments (MoM) equation is $\overline{Y_n} = (1/n) \sum_{i=1}^{n} Y_i = \mathrm{E}_\theta(Y)$.

The second equation is $(1/n) \sum_{i=1}^{n} Y_i^2 = \mathrm{E}_\theta(Y^2)$. The third equation is $(1/n) \sum_{i=1}^{n} Y_i^3 = \mathrm{E}_\theta(Y^3)$.....

The equation(s) are solved to give an estimator of θ, say W. Then the estimator of $h(\theta)$ is $h(W)$.

7. Properties of estimators

(a) *Bias* of estimator T of θ is $b_T(\theta) \; = \; \mathrm{E}_\theta(T) - \theta$.

(b) *MSE* of estimator T of θ is $mse_\theta(T) \; = \; \mathrm{E}((T-\theta)^2) \; = \; \mathrm{var}_\theta(T) + (b_T(\theta))^2$.

(c) An estimator T_n based on an n-sample is *asymptotically unbiased* if $b_{T_n}(\theta) \to \; 0$ as $n \to \; \infty$.

(d) An estimator T_n based on an n-sample is *consistent* if, for all $\epsilon > 0$, $P(|T_n - \theta| \; > \; \epsilon) \; \to \; 0$ as $n \to \; \infty$.

(e) Chebychev's inequality: consistency and MSE are related by the fact that, for all $\epsilon > 0$,

$$P(|T_n - \theta| \; > \; \epsilon) \; \leq \; \mathrm{E}((T_n - \theta)^2)/\epsilon^2.$$

8. Likelihood, MLE, and sufficient statistics (for an n-sample $Y_1, ..., Y_n$ from $f_Y(y;\theta)$)

(a) The *likelihood* $L(\theta)$ is the joint pdf/pmf of $Y_1, ..., Y_n$, evaluated at the observed data $y_1, ..., y_n$.

(b) That is $L_n(\theta) = \prod_{i=1}^n f_Y(y_i;\theta)$, and the log-likelihood $\ell_n(\theta) \; = \; \log_e L_n(\theta) \; = \; \sum_{i=1}^n \log f_Y(y_i;\theta)$.

(c) The maximum likelihood estimate (MLE) is the value of θ that maximizes the likelihood or log-likelihood.

(d) Statistic T is *sufficient*, if the conditional pdf/pmf of $Y_1, ..., Y_n$ given $T = t$ does not depend on θ.

(e) **Factorization criterion:** T is sufficient if and only if $L_n(\theta) = b(y_1, ..., y_n).g(t;\theta)$ for $b()$ not involving θ and $g()$ involving $(y_1, ..., y_n)$ only through $T(y_1, ..., y_n) = t$.

(f) **Rao-Blackwell Theorem:** if an estimator W is not a function of the sufficient statistic, T, then there is an estimator which is a function of T with same expectation as W and smaller mean square error.

9. The Cramér-Rao Lower Bound

(a) Let $Y_1, ..., Y_n$ be n-sample from pmf or pdf $f_Y(y;\theta)$, where $f_Y()$ can be differentiated at least twice w.r.t θ. Let T be any unbiased estimator of θ, based on $Y_1, ..., Y_n$. Then

$$\mathrm{var}(T) \quad \geq \quad \left\{ n\mathrm{E}\left[\left(\frac{\partial \log_e f_Y(Y;\theta)}{\partial \theta} \right)^2 \right] \right\}^{-1} \quad = \quad \left\{ -n\mathrm{E}\left[\left(\frac{\partial^2 \log_e f_Y(Y;\theta)}{\partial \theta^2} \right) \right] \right\}^{-1}$$

(b) This lower bound on the variance of an unbiased estimator is the *Cramér-Rao Lower Bound* (CRLB).

(c) The *efficiency* of an unbiased estimator T is the ratio of the CRLB to the the variance $\mathrm{var}(T)$.

(d) If T is unbiased and $\mathrm{var}(T)$ is equal to the CRLB, then T is *efficient*.

10. Interval Estimation

(a) Let $Y_1, ..., Y_n$ be an n-sample from some pmf or pdf indexed by parameter θ. Let $L(Y_1, ..., Y_n)$ and $U(Y_1, ..., Y_n)$ be two functions of the data random variables, such that $L \leq U$ for all possible samples $y_1, ..., y_n$.

If $$P_\theta(L(Y_1, ..., Y_n) \leq \theta \leq U(Y_1, ..., Y_n)) \; = \; (1-\alpha) \text{ for all } \theta$$

then $(L(y_1, ..., y_n), \; U(y_1, ..., y_n))$ is a $(1-\alpha)$-level confidence interval for θ.

(b) Confidence intervals should be based on sufficient statistics. If T is sufficient, we find $L(T)$ and $U(T)$ such that $L(t) \leq U(t)$ for all values t of T, and $P(L(T) \leq \theta \leq U(T)) \; = \; (1-\alpha)$.

(c) Conventionally, we choose $L(t)$ and $U(t)$ such that $P(L(T) \; > \; \theta) \; = \; P(U(T) \; < \; \theta) \; = \; \alpha/2$.

11. Bayesian Estimation

(a) In Bayesian inference a *prior* pdf $\pi(\theta)$ is assigned to θ.

(b) Given an n-sample $Y_1, ..., Y_n$ from $f_Y(y;\theta)$, and sufficient statistic T, the *posterior* pdf of θ given the sample values $y_1, ..., y_n$ is the same as the posterior pdf given the value t of T.

(c) The posterior pdf of θ is $\pi(\theta \mid T = t) \; = \; f_T(t;\theta)\pi(\theta) / \int_\theta f_T(t;\theta)\pi(\theta) \; d\theta$.

(d) All estimates are based on the posterior pdf; this may be a point estimate such as the mean of the posterior pdf, or an interval estimate $(P(L(t) < \theta < U(t) \mid T = t) \; = \; 1 - \alpha)$.

12. Standard distributions:

	pmf or pdf	mean	variance
(a) Binomial; $B(n,p)$	$P(X=k) = \binom{n}{k} p^k(1-p)^{n-k}$	np	$np(1-p)$
index n, parameter p	$k = 0,1,2,...,n$		
(b) Geometric; $Geo(p)$; parameter p	$P(X=k) = p(1-p)^{k-1}$, $k = 1,2,3,4,......$	$1/p$	$(1-p)/p^2$
(c) Neg. Binomial; $NegB(r,p)$;	$P(X=k) = \binom{k-1}{r-1} p^r(1-p)^{k-r}$	r/p	$r(1-p)/p^2$
index r, parameter p	$k = r, r+1, r+2, ...$		
(d) Poisson; $\mathcal{P}o(\mu)$	$P(X=k) = \exp(-\mu)\mu^k/k!, \quad k = 0,1,2,...$	μ	μ
(e) Uniform on (a,b); $U(a,b)$;	$f_X(x) = 1/(b-a), \quad a < x < b$	$(b+a)/2$	$(b-a)^2/12$
(f) Normal, $N(\mu,\sigma^2)$	$f_X(x) = (1/\sqrt{2\pi\sigma^2})\exp(-(x-\mu)^2/2\sigma^2)$	μ	σ^2
(g) Exponential, $\mathcal{E}(\lambda)$	$f_X(x) = \lambda\exp(-\lambda x)$	$1/\lambda$	$1/\lambda^2$
rate parameter λ	$0 \leq x < \infty$		
(h) Gamma, $G(\alpha,\lambda)$	$f_X(x) = \lambda^\alpha x^{\alpha-1}\exp(-\lambda x)/\Gamma(\alpha)$	α/λ	α/λ^2
shape α, rate λ	$0 \leq x < \infty$		
(j) Beta $Be(r,s)$	$f_X(x) = \frac{\Gamma(r+s}{\Gamma(r)\Gamma(s)} x^{r-1}(1-x)^{s-1}$	$r/(r+s)$	
$r > 0$,$s > 0$	$0 \leq x \leq 1$		

Note: $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\exp(-x)dx \qquad \Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$, and $\Gamma(n) = (n-1)!$ for integer n.

13. Moment generating functions: $M_X(t) = E(\exp(tX))$

	mgf	note
(a) Binomial; $B(n,p)$	$(q + p\ z))^n$	where $q = 1-p$ and $z \equiv \exp(t)$
(b) Geometric; $Geo(p)$;	$pz/(1-q\ z)$	where $q = 1-p$ and $z \equiv \exp(t)$
(c) Neg. Binomial; $NegB(r,p)$;	$(pz/(1-q\ z))^r$	where $q = 1-p$ and $z \equiv \exp(t)$
(d) Poisson; $\mathcal{P}o(\mu)$	$\exp(\mu(z-1))$	where $z \equiv \exp(t)$
(g) Exponential, $\mathcal{E}(\lambda)$	$\lambda/(\lambda-t)$	provided $t < \lambda$
(h) Gamma, $G(\alpha,\lambda)$	$(\lambda/(\lambda-t))^\alpha$	provided $t < \lambda$
(i) Z^2 where $Z \sim N(0,1)$	$(\frac{1}{2}/(\frac{1}{2}-t))^{1/2}$	provided $t < \frac{1}{2}$

On this page will be Normal and Chi-squared tables copied from Pages 852 and 856 in the text book