

Probability and Statistics

1. There are r players, with player i initially having n_i units, $n_i > 0$, $i = 1, \dots, r$. At each stage, two of the players are chosen to play a game, with the winner of the game receiving 1 unit from the loser. Any player whose fortune drops to 0 is eliminated, and this continues until a single player has all $n = \sum_{i=1}^r n_i$ units, with that player designated as the winner. Note that the mechanism to choose two players at each stage is unknown. It can be either deterministic or random. Assume that the results of successive games are independent and that each game is equally likely to be won by either of its two players.

For any set of players $S \subseteq \{1, \dots, r\}$, let $X(S)$ denote the number of games involving only members of S . Does $E(X(S))$ depend on the player selection mechanism? If you think it doesn't depend, calculate the expectation. If you think it depends, give two mechanisms leading to different expectations.

2. Let X_1, X_2, \dots be independent Bernoulli random variables satisfying $P(X_i = 1) = p$ and $P(X_i = -1) = q = 1 - p$ for some $p \in (0, 1)$. Let $S_n = X_1 + \dots + X_n$ and $M = \sup_{n \geq 1} (S_n/n)$.
 - (a) Calculate $P(M = 0)$.
 - (b) Show that $P(p - q < M \leq 1) = 1$. For any rational number $x \in (p - q, 1]$, is $P(M = x) > 0$? If so, prove it. If not, find a point with zero probability.
3. Let X have a uniform distribution on the interval $[0, 1]$ and let $N_{m,k}$ be the digit in the m th place to the right of the decimal point in X^k .
 - (a) Find $\lim_{m \rightarrow \infty} P(N_{m,m} = i)$ for $i = 0, 1, 2, \dots, 9$.
 - (b) Let $k(m)$ be a function of m , taking values greater than 1. Find a necessary and sufficient condition on $k(m)$ such that $\lim_{m \rightarrow \infty} P(N_{m,k(m)} = i) = \frac{1}{10}$ for $i = 0, 1, 2, \dots, 9$.

4. Assume we have n observations: (Y_i, \mathbf{x}_i) , $i = 1, \dots, n$, where Y_i is the random response and $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ is a vector of p fixed covariates for the i th observation. Denote $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ be a unknown p -length vector of regression coefficients. Let $\theta_i = \sum_{j=1}^p x_{ij}\beta_j$, $\mu_i = E(Y_i)$ and $\sigma_i^2 = \text{Var}(Y_i)$. Assume the density of Y_i belongs to the following exponential family:

$$f(y_i; \theta_i) = \exp\{\theta_i y_i - b(\theta_i)\}, \quad (1)$$

where $b'(\theta_i) = \mu_i$, $b''(\theta_i) = \sigma_i^2$. Suppose that all θ_i 's are contained in a compact subset of a space Θ . Let $\ell_n(\boldsymbol{\beta})$ be the log-likelihood function of the data, and let $\mathbf{H}_n(\boldsymbol{\beta}) = -\frac{\partial^2 \ell_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}$.

Let \mathcal{X} be the set of all p covariates under consideration. Let $\alpha_0 \subset \mathcal{X}$ be the subset that contains and only contains all the important covariates affecting Y (the corresponding β_j 's are nonzero). Let α be any subset of \mathcal{X} , and let $\beta(\alpha)$ be the vector of the components in β that correspond to the covariates in α . Let $A = \{\alpha : \alpha_0 \subset \alpha\}$ be the collection of models that including all important covariates. We assume:

- (I) There exist positive constants C_1, C_2 such that for all sufficiently large n ,

$$C_1 < \lambda_{\min} \left\{ \frac{1}{n} H_n(\beta) \right\} < \lambda_{\max} \left\{ \frac{1}{n} H_n(\beta) \right\} < C_2,$$

where $\lambda_{\min} \left\{ \frac{1}{n} H_n(\beta) \right\}$ and $\lambda_{\max} \left\{ \frac{1}{n} H_n(\beta) \right\}$ are the smallest and largest eigenvalues of $\frac{1}{n} H_n(\beta)$.

- (II) For any given $\epsilon > 0$, there exists a constant $\delta > 0$ such that, when n is sufficiently large,

$$(1 - \epsilon) H_n(\beta(\alpha)) \leq H_n(\tilde{\beta}) \leq (1 + \epsilon) H_n(\beta(\alpha))$$

for all $\alpha \in A$ and $\tilde{\beta}$ satisfying $\|\tilde{\beta} - \beta(\alpha)\| \leq \delta$.

For any model α , let $\hat{\beta}_\alpha$ be the MLE of $\beta(\alpha)$ based on this model. Show that

$$\max_{\alpha \in A} \|\hat{\beta}_\alpha - \beta(\alpha)\| = O_p(n^{-1/3}).$$

5. Consider a random sample of size n , and write the data as an $r = r_n$ by $c = c_n$ matrix, $\{X_{ij} : i = 1, \dots, r_n; j = 1, \dots, c_n\}$ with $n = r_n c_n$. To specify notation, $\{X_{ij}\}$ are i.i.d. with c.d.f. $F(x)$ and continuous density $f(x)$. Let β denote the median, i.e., $F(\beta) = 0.5$. Define an estimator by

$$\hat{\beta}_n = \min_j \left\{ \max_i \{X_{ij}\} \right\}.$$

- (a) What is the condition on r_n when $n \rightarrow \infty$ for median-unbiasedness, i.e., β is also the median for the distribution of $\hat{\beta}_n$?
- (b) We further assume F is differentiable in an open neighborhood of β and has a positive derivative at β . For r_n in (a), show that $r_n(\hat{\beta}_n - \beta)$ converges in distribution, and find the limiting distribution function.