

# Midterm Formula Sheet

## Contents

<b>1 Exam formula sheet</b>	<b>1</b>
1.1 Epi 202: Probability . . . . .	1
1.2 Epi 203: Statistical inference . . . . .	1
1.3 Epi 204: Generalized linear models . . . . .	2
1.3.1 Estimates of odds ratios from 2x2 contingency tables . . . . .	2
1.3.2 Survival analysis . . . . .	3

## 1 Exam formula sheet

### 1.1 Epi 202: Probability

$$\begin{aligned}\text{Var}(\tilde{a} \cdot \tilde{X}) &= \text{Var}\left(\sum_{i=1}^n a_i X_i\right) \\ &= \tilde{a}^\top \text{Var}(\tilde{X}) \tilde{a} \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(X_i, X_j)\end{aligned}$$

### 1.2 Epi 203: Statistical inference

$$\mathcal{L}(\theta) \stackrel{\text{def}}{=} \text{p}(\tilde{X} = \tilde{x} | \Theta = \theta)$$

$$\ell \stackrel{\text{def}}{=} \log\{\mathcal{L}(\tilde{x}|\theta)\}$$

$$\ell' \stackrel{\text{def}}{=} \frac{\partial}{\partial \theta} \ell(\tilde{x}|\theta)$$

$$\ell'' \stackrel{\text{def}}{=} \frac{\partial}{\partial \tilde{\theta}} \frac{\partial}{\partial \tilde{\theta}^\top} \ell(\tilde{x}|\tilde{\theta})$$

$$\ell''_{ij} = \frac{\partial}{\partial \theta_i} \frac{\partial}{\partial \theta_j} \ell(\tilde{X} = \tilde{x}|\tilde{\theta})$$

$$I \stackrel{\text{def}}{=} -\ell''(\tilde{x}|\tilde{\theta})$$

$$\mathcal{J} \stackrel{\text{def}}{=} \text{E}[I(\tilde{x}|\theta)]$$

$$\hat{\theta}_{ML} \sim \text{N}\left(\theta, [\mathcal{J}(\tilde{\theta})]^{-1}\right)$$

### 1.3 Epi 204: Generalized linear models

Generalized linear models have three components:

1. The **outcome distribution** family:  $p(Y|\mu(\tilde{x}))$
2. The **link function**:  $g(\mu(\tilde{x})) = \eta(\tilde{x})$
3. The **linear component**:  $\eta(\tilde{x}) = \tilde{x} \cdot \beta$

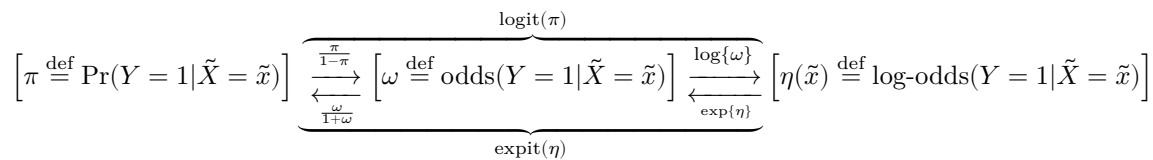


Figure 1: Diagram of logistic regression link and inverse link functions

$$\theta(\tilde{x}, \tilde{x}^*) = \exp\{(\Delta\tilde{x}) \cdot \tilde{\beta}\}$$

#### 1.3.1 Estimates of odds ratios from 2x2 contingency tables

$$\hat{\theta} = \frac{ad}{bc}$$

### 1.3.2 Survival analysis

#### Probability distribution functions

Table 1: Probability distribution functions

Name	Symbols	Definition
Probability density function (PDF)	$f(t), p(t)$	$p(T = t)$
Cumulative distribution function (CDF)	$F(t), P(t)$	$P(T \leq t)$
Survival function	$S(t), \bar{F}(t)$	$P(T > t)$
Hazard function	$\lambda(t), h(t)$	$p(T = t   T \geq t)$
Cumulative hazard function	$\Lambda(t), H(t)$	$\int_{u=-\infty}^t \lambda(u) du$
Log-hazard function	$\eta(t)$	$\log\{\lambda(t)\}$

#### Diagram of survival distribution function relationships

$$\begin{aligned}
 f(t) &\xleftarrow[\frac{-S'(t)}{S(t)\lambda(t)}]{} S(t) \xleftarrow[\exp\{-\Lambda(t)\}]{} \Lambda(t) \xleftarrow[\int_{u=0}^t \lambda(u) du]{} \lambda(t) \xleftarrow[\exp\{\eta(t)\}]{} \eta(t) \\
 f(t) &\xrightarrow[\int_{u=t}^{\infty} f(u) du]{} S(t) \xrightarrow[-\log S(t)]{} \Lambda(t) \xrightarrow[\Lambda'(t)]{} \lambda(t) \xrightarrow[\log\{\lambda(t)\}]{} \eta(t)
 \end{aligned}$$

#### Survival likelihood contributions, assuming non-informative censoring

$$\begin{aligned}
 p(Y = y, D = d) &= [f_T(y)]^d [S_T(y)]^{1-d} \\
 &= [\lambda_T(y)]^d [S_T(y)]
 \end{aligned}$$

#### Nonparametric time-to-event distribution estimators

$$\hat{\lambda}_i = \frac{d_i}{n_i}$$

$$\hat{S}_{KM}(t) \stackrel{\text{def}}{=} \prod_{\{i: t_i < t\}} [1 - \hat{\lambda}_i]$$

$$\hat{\Lambda}_{NA}(t) \stackrel{\text{def}}{=} \sum_{\{i: t_i < t\}} \hat{\lambda}_i$$

## Proportional hazards model structure

**Joint likelihood of data set:**  $\mathcal{L} \stackrel{\text{def}}{=} p(\tilde{Y} = \tilde{y}, \tilde{D} = \tilde{d} | \mathbf{X} = \mathbf{x})$

**Marginal likelihood contribution of obs.  $i$ :**  $\mathcal{L}_i \stackrel{\text{def}}{=} p(Y_i = y_i, D_i = d_i | \tilde{X}_i = \tilde{x}_i)$

*Independent Observations Assumption:*  $\mathcal{L} = \prod_{i=1}^n \mathcal{L}_i$

*Non-Informative Censoring Assumption:*  $T_i \perp\!\!\!\perp C_i | \tilde{X}_i$

$$\mathcal{L}_i \propto [f_T(y_i | \tilde{x}_i)]^{d_i} [S_T(y_i | \tilde{x}_i)]^{1-d_i} = S_T(y_i | \tilde{x}_i) \cdot [\lambda_T(y_i | \tilde{x}_i)]^{d_i}$$

**Survival function:**  $S(t | \tilde{x}) \stackrel{\text{def}}{=} P(T > t | \tilde{X} = \tilde{x}) = \int_{u=t}^{\infty} f(u | \tilde{x}) du = \exp\{-\Lambda(t | \tilde{x})\}$

**Probability density function:**  $f(t | \tilde{x}) \stackrel{\text{def}}{=} p(T = t | \tilde{X} = \tilde{x}) = -S'(t | \tilde{x}) = \lambda(t | \tilde{x}) S(t | \tilde{x})$

**Cumulative hazard function:**  $\Lambda(t | \tilde{x}) \stackrel{\text{def}}{=} \int_{u=0}^t \lambda(u | \tilde{x}) du = -\log\{S(t | \tilde{x})\}$

**Hazard function:**  $\lambda(t | \tilde{x}) \stackrel{\text{def}}{=} p(T = t | T \geq t, \tilde{X} = \tilde{x}) = \Lambda'(t | \tilde{x}) = \frac{f(t | \tilde{x})}{S(t | \tilde{x})}$

**Hazard ratio:**  $\theta(t | \tilde{x} : \tilde{x}^*) \stackrel{\text{def}}{=} \frac{\lambda(t | \tilde{x})}{\lambda(t | \tilde{x}^*)}$

**Log-Hazard function:**  $\eta(t | \tilde{x}) \stackrel{\text{def}}{=} \log\{\lambda(t | \tilde{x})\} = \eta_0(t) + \Delta\eta(t | \tilde{x})$

*Proportional Hazards Assumption:*

$$\begin{aligned}\lambda(t | \tilde{x}) &= \lambda_0(t) \cdot \theta(\tilde{x}) \\ \Lambda(t | \tilde{x}) &= \Lambda_0(t) \cdot \theta(\tilde{x}) \\ \eta(t | \tilde{x}) &= \eta_0(t) + \Delta\eta(\tilde{x})\end{aligned}$$

*Logarithmic Link Function Assumption:*

- **Link function:**

$$\begin{aligned}\log\{\lambda(t | \tilde{x})\} &= \eta(t | \tilde{x}) \\ \log\{\theta(\tilde{x})\} &= \Delta\eta(\tilde{x})\end{aligned}$$

- **Inverse link function:**

$$\begin{aligned}\lambda(t | \tilde{x}) &= \exp\{\eta(t | \tilde{x})\} \\ \theta(\tilde{x}) &= \exp\{\Delta\eta(\tilde{x})\}\end{aligned}$$

**Linear Predictor Component:**

$$\begin{aligned}\eta(t | \tilde{x}) &= \eta_0(t) + \Delta\eta(t | \tilde{x}) \\ \Delta\eta(t | \tilde{x}) &= \tilde{x} \cdot \tilde{\beta}\end{aligned}$$

*Linear Predictor Component Functional Form Assumption:*

$$\Delta\eta(t | \tilde{x}) = \tilde{x} \cdot \tilde{\beta} \stackrel{\text{def}}{=} \beta_1 x_1 + \dots + \beta_p x_p$$

**Proportional hazards model partial likelihood formula:**

$$\mathcal{L}_i^* = \frac{\theta(\tilde{x}_i)}{\sum_{k \in R(t_i)} \theta(\tilde{x}_k)}$$
$$\mathcal{L}^* = \prod_{\{i: d_i=1\}} \mathcal{L}_i^*$$

**Proportional hazards model baseline cumulative hazard estimator:**

$$\hat{\Lambda}_0(t) = \sum_{t_i < t} \frac{d_i}{\sum_{k \in R(t_i)} \theta(x_k)}$$