# Midterm Formula Sheet

## Contents



# 1 Exam formula sheet

## 1.1 Epi 202: Probability

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

## 1.2 Epi 203: Statistical inference

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

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

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

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

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

$$\begin{aligned} I &\stackrel{\text{def}}{=} -\ell''(\tilde{x}|\tilde{\theta}) \\ \mathcal{I} &\stackrel{\text{def}}{=} \mathrm{E}[I(\tilde{x}|\theta)]\end{aligned}$$

$$\hat{\theta}_{ML} \mathrel{\dot{\sim}} \mathrm{N}\left(\theta, \left[\mathcal{I}(\tilde{\theta})\right]^{-1}\right)$$

## 1.3 Epi 204: Generalized linear models

Generalized linear models have three components:

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

$$\left[\pi \stackrel{\text{def}}{=} \Pr(Y=1|\tilde{X}=\tilde{x})\right] \overbrace{\underbrace{\underset{\xleftarrow[\frac{\omega}{1+\omega}]{}}{\xrightarrow{\frac{\pi}{1-\pi}}} \left[\omega \stackrel{\text{def}}{=} \text{odds}(Y=1|\tilde{X}=\tilde{x})\right] \underset{\xleftarrow[\exp\{\eta\}]{}}{\xrightarrow{\log\{\omega\}}}}_{\text{expit}(\eta)}}^{\text{logit}(\pi)} \left[\eta(\tilde{x}) \stackrel{\text{def}}{=} \text{log-odds}(Y=1|\tilde{X}=\tilde{x})\right]$$

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

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

### 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) | $\mathrm{f}(t), \mathrm{p}(t)$ | $\mathrm{p}(T = t)$ |
| Cumulative distribution function (CDF) | $\mathrm{F}(t), \mathrm{P}(t)$ | $\mathrm{P}(T \leq t)$ |
| Survival function | $\mathrm{S}(t), \bar{\mathrm{F}}(t)$ | $\mathrm{P}(T > t)$ |
| Hazard function | $\lambda(t), \mathrm{h}(t)$ | $\mathrm{p}(T = t \mid T \geq t)$ |
| Cumulative hazard function | $\Lambda(t), \mathrm{H}(t)$ | $\int_{u=-\infty}^{t} \lambda(u) du$ |
| Log-hazard function | $\eta(t)$ | $\log\{\lambda(t)\}$ |

#### Diagram of survival distribution function relationships

$$\mathrm{f}(t) \xleftarrow[\mathrm{S}(t)\lambda(t)]{-S'(t)} \mathrm{S}(t) \xleftarrow{\exp\{-\Lambda(t)\}} \Lambda(t) \xleftarrow{\int_{u=0}^{t} \lambda(u) du} \lambda(t) \xleftarrow{\exp\{\eta(t)\}} \eta(t)$$

$$\mathrm{f}(t) \xrightarrow[\int_{u=t}^{\infty} \mathrm{f}(u) du]{\mathrm{f}(t)/\lambda(t)} \mathrm{S}(t) \xrightarrow[-\log \mathrm{S}(t)]{} \Lambda(t) \xrightarrow[\Lambda'(t)]{} \lambda(t) \xrightarrow[\log\{\lambda(t)\}]{} \eta(t)$$

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

$$\begin{aligned}
\mathrm{p}(Y = y, D = d) &= [\mathrm{f}_T(y)]^d [\mathrm{S}_T(y)]^{1-d} \\
&= [\lambda_T(y)]^d [\mathrm{S}_T(y)]
\end{aligned}$$

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

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

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

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

**Proportional hazards model structure**

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

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

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

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

$$\mathcal{L}_i \propto [\text{f}_T(y_i|\tilde{x}_i)]^{d_i}[\text{S}_T(y_i|\tilde{x}_i)]^{1-d_i} = \text{S}_T(y_i|\tilde{x}_i) \cdot [\lambda_T(y_i|\tilde{x}_i)]^{d_i}$$

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

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

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

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

**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**:

$$\log\{\lambda(t|\tilde{x})\} = \eta(t|\tilde{x})$$

$$\log\{\theta(\tilde{x})\} = \Delta\eta(\tilde{x})$$

- **Inverse link function**:

$$\lambda(t|\tilde{x}) = \exp\{\eta(t|\tilde{x})\}$$

$$\theta(\tilde{x}) = \exp\{\Delta\eta(\tilde{x})\}$$

**Linear Predictor Component**:

$$\eta(t|\tilde{x}) = \eta_0(t) + \Delta\eta(t|\tilde{x})$$

$$\Delta\eta(t|\tilde{x}) = \tilde{x} \cdot \tilde{\beta}$$

*Linear Predictor Component Functional Form Assumption:*

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

**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:**

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