



Evaluation of DNA Mixtures Accounting for Sampling Variability

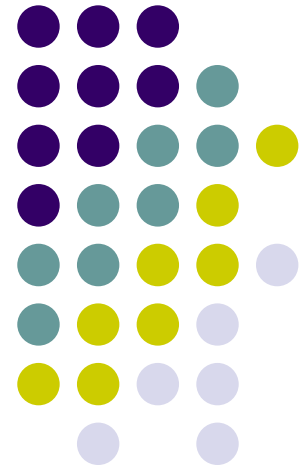
Yuk-Ka Chung

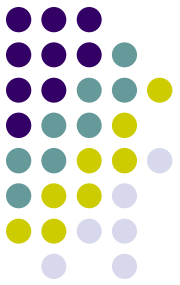
Dept. of Statistics and Actuarial Science

University of Hong Kong

yukchung@hku.hk

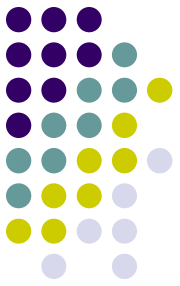
Joint work: Y. Q. Hu, D. G. Zhu, W. K. Fung





Outline

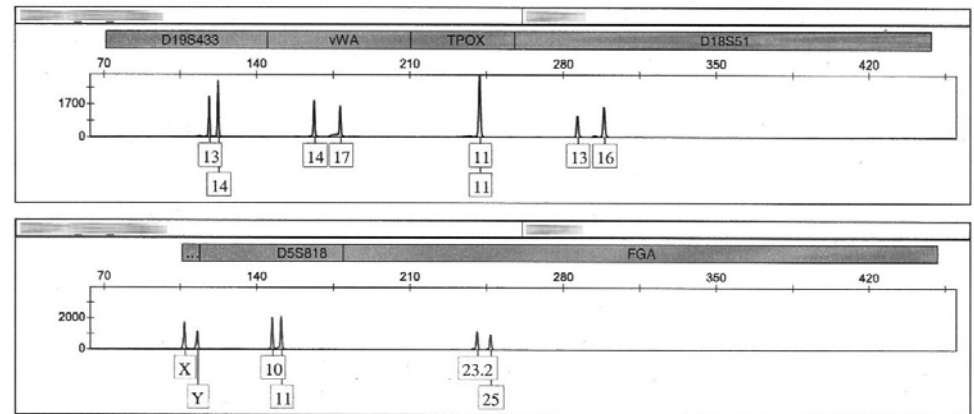
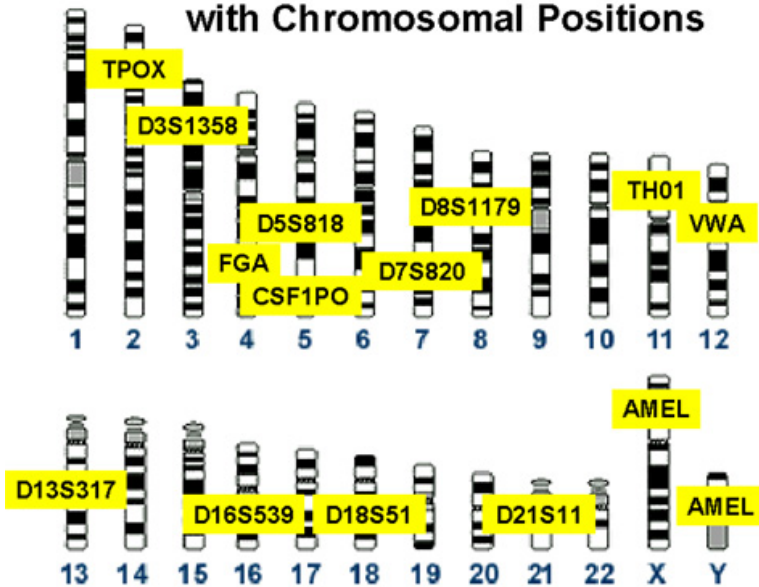
- Background
- Bayesian model
- Example
- Conclusion and future works



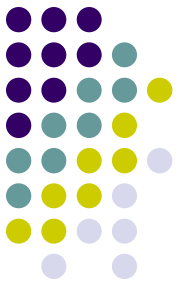
DNA Profile

- Set of numbers representing the genetic characteristics of an individual

13 CODIS Core STR Loci with Chromosomal Positions



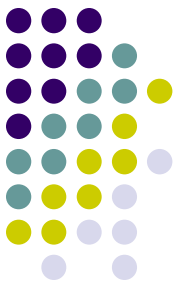
D19S433	vWA	TPOX
13 / 14	14 / 17	11 / 11
D18S51	D5S818	FGA
13 / 16	10 / 11	23.2 / 25



Two-person Mixture

- Crime stain M
- Reference samples K
= (V, S)
 - Victim V , Suspect S
- Evidentiary value

$$LR = \frac{P(M, K | H_p)}{P(M, K | H_d)} = \frac{P(M | K, H_p)}{P(M | K, H_d)}$$



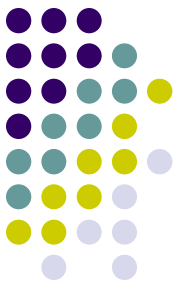
Two-person Mixture

Rape case in HK

Locus	Mixture	Victim	Suspect	Frequency
D5S818	7		7	0.035
	10	10		0.252
	13	13		0.165

H_p : the victim and the suspect were contributors

$$P(M | K, H_p) = 1$$



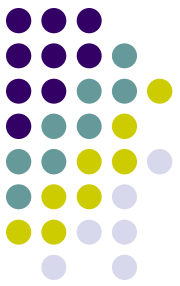
Two-person Mixture

Rape case in HK

Locus	Mixture	Victim	Suspect	Frequency
D5S818	7		7	0.035
	10	10		0.252
	13	13		0.165

H_d : the victim and **one unknown** were contributors

$$P(M | K, H_d) = p_7^2 + 2p_7p_{10} + 2p_7p_{13} = 0.0306$$



Two-person Mixture

Rape case in HK

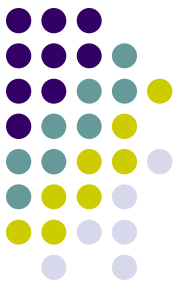
Locus	Mixture	Victim	Suspect	Frequency
D5S818	7		7	0.035
	10	10		0.252
	13	13		0.165

} Estimated
by
database D

H_p : the victim and the suspect were contributors

H_d : the victim and **one unknown** were contributors

$$LR = \frac{1}{0.0306} = 32.7$$



Two-person Mixture

Rape case in HK

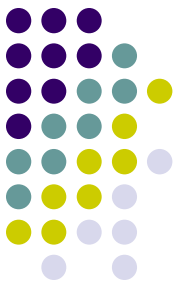
Locus	Mixture	Victim	Suspect	Frequency
D5S818	7		7	0.050
	10	10		0.252
	13	13		0.165

H_p : the victim and the suspect were contributors

H_d : the victim and **one unknown** were contributors

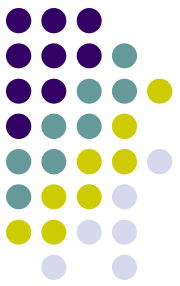
$$LR = \frac{1}{0.0442} = 22.6$$

31% reduction



Sampling Variability

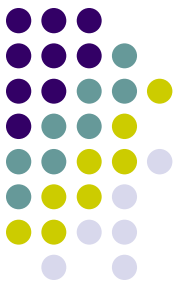
- Underestimate the allele frequencies
- “Overstate the strength of the evidence against the defendant” (Balding, 1995)
- Bayesian approaches on identification cases
 - Balding (1995)
 - Balding & Donnelly (1995)
 - Foreman *et al.* (1997)
 - Curran *et al.* (2002)
 - Corradi *et al.* (2003)



Bayesian Model

- Database D with n individuals
- Reference samples (K, D)
- Evidential value

$$LR = \frac{P(M, K, D | H_p)}{P(M, K, D | H_d)}$$



Bayesian Model

- Allele frequencies

$$x_{\theta} = (x_{\theta,1}, x_{\theta,2}, \dots, x_{\theta,l})$$

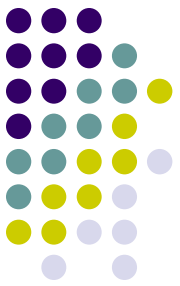
- Dirichlet prior

$$Dir(x_{\theta} | \alpha)$$

$$\begin{aligned} P(M, K, D | H) &= \int_{x_{\theta}} P(M, K, D | x_{\theta}, H) Dir(x_{\theta} | \alpha) dx_{\theta} \\ &= \int_{x_{\theta}} P(M | x_{\theta}, K, H) P(K, D | x_{\theta}) Dir(x_{\theta} | \alpha) dx_{\theta} \end{aligned}$$

Probability by “plug-in” approach

Product multinomial



Bayesian Model

Rape case in HK

Locus	Mixture	Victim	Suspect	Frequency in D
D5S818	7		7	n_7
	10	10		n_{10}
	13	13		n_{13}

H_p : the victim and the suspect were contributors

H_d : the victim and one unknown were contributors

$$\alpha = (1 \quad 1 \quad \dots \quad 1)'$$

$$LR = \frac{(12 + 2n)(13 + 2n)}{(3 + n_7)(12 + n_7 + 2n_{10} + 2n_{13})}$$



Bayesian Model

Rape case in HK

Locus	Mixture	Victim	Suspect	Frequency in D
D5S818	7		7	20
	10	10		143
	13	13		94

$n = 284$

H_p : the victim and the suspect were contributors

H_d : the victim and one unknown were contributors

$$LR = 28.96$$



Bayesian Model

Rape case in HK

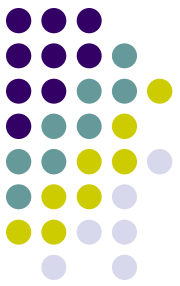
Locus	Mixture	Victim	Suspect	Frequency in D
D5S818	7		7	70
	10	10		504
	13	13		331

$n = 1000$

H_p : the victim and the suspect were contributors

H_d : the victim and one unknown were contributors

$$LR = 31.58 \longrightarrow 32.7$$



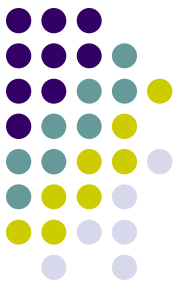
Q-function

- “plug-in” probability $P(M \mid x_\theta, K, H)$ as a linear combination of the Q-function (Fung & Hu, 2008)

$$Q(j, B \mid x_\theta) = \sum_{M \setminus B \subset C \subset M} (-1)^{|M \setminus C|} \left(\sum_{i \in C} x_{\theta, i} \right)^j$$

- Probability of j random alleles that all belong to M and explain all alleles in B

$$P(B \subset X \subset M)$$



Q-function

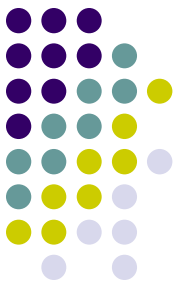
- $M = \{ 7, 10, 13 \}$, $V = \{ 10, 13 \}$, $S = \{ 7, 7 \}$
- H_d : V and one unknown were contributors

$$P(M \mid x_\theta, K, H_d) = Q(2, \{ 7 \})$$

- $H_{d'}$: V and one **relative of S** were contributors

$$P(M \mid x_\theta, K, H_{d'}) = k_0 Q(2, \{ 7 \}) + 2k_1 Q(1, \phi) + k_2 Q(0, \phi)$$

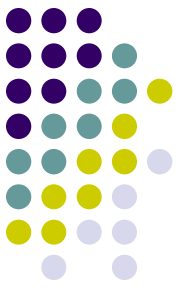
k_i : Kinship coefficients



Q-function

- Suspect is typed but prosecution involves a close relative
- Suspect may be unavailable for typing and a close relative is typed

H : the victim, one relative R of a typed person T and other $x - 1$ unknowns were contributors

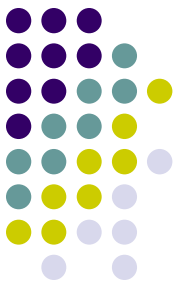


Q-function

H : the victim, one relative R of $T = t_1 t_2$ and other $x - 1$ unknowns were contributors

U — set of alleles in M but absent in known contributors declared in H

$$\begin{aligned} P(M \mid x_\theta, K, H) = & k_0 Q(2x, U) + k_1 I_M(t_1) Q(2x - 1, U \setminus \{t_1\}) \\ & + k_1 I_M(t_2) Q(2x - 1, U \setminus \{t_2\}) \\ & + k_2 I_M(t_1) I_M(t_2) Q(2x - 2, U \setminus \{t_1, t_2\}) \end{aligned}$$

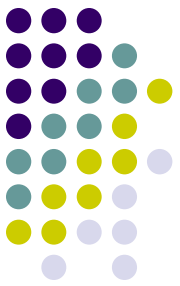


Modified Q-function

$$P(M, K, D | H) = \int_{\chi_\theta} P(M | x_\theta, K, H) P(K, D | x_\theta) \text{Dir}(x_\theta | \alpha) dx_\theta$$

$$Q^*(j, B) = \int_{\chi_\theta} Q(j, B | x_\theta) P(K, D | x_\theta) \text{Dir}(x_\theta | \alpha) dx_\theta$$

- Replacing $Q(.,. | x_\theta)$ in $P(M | x_\theta, K, H)$ by $Q^*(.,.)$ gives $P(M, D, K | H)$



Modified Q-function

$$Q^*(j, B) \propto \frac{\Gamma(\alpha_{\bullet} + 2k + 2n)}{\Gamma(\alpha_{\bullet} + 2k + 2n + j)} \sum_{M \setminus B \subset C \subset M} (-1)^{|M \setminus C|} \frac{\Gamma(\alpha_C + n_{C, K \cup D} + j)}{\Gamma(\alpha_C + n_{C, K \cup D})}$$

$$\frac{\Gamma(r + j)}{\Gamma(r)} = r(r + 1) \cdots (r + j - 1)$$

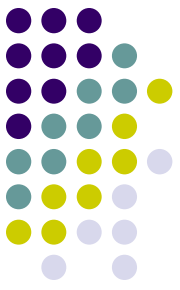
- Closed-form formula
- Easy to be implemented by computer program
- No simulation or approximation is need
- Does not increase computational complexity



Numerical Example

Rape case in HK

Locus	Mixture	Victim	Suspect	Frequency
FGA	18		18	0.025
	19		19	0.065
	24	24		0.166
	26	26		0.048
D5S818	7		7	0.035
	10	10		0.252
	13	13		0.165
D8S1179	12	12		0.118
	16		16	0.098



Numerical Example

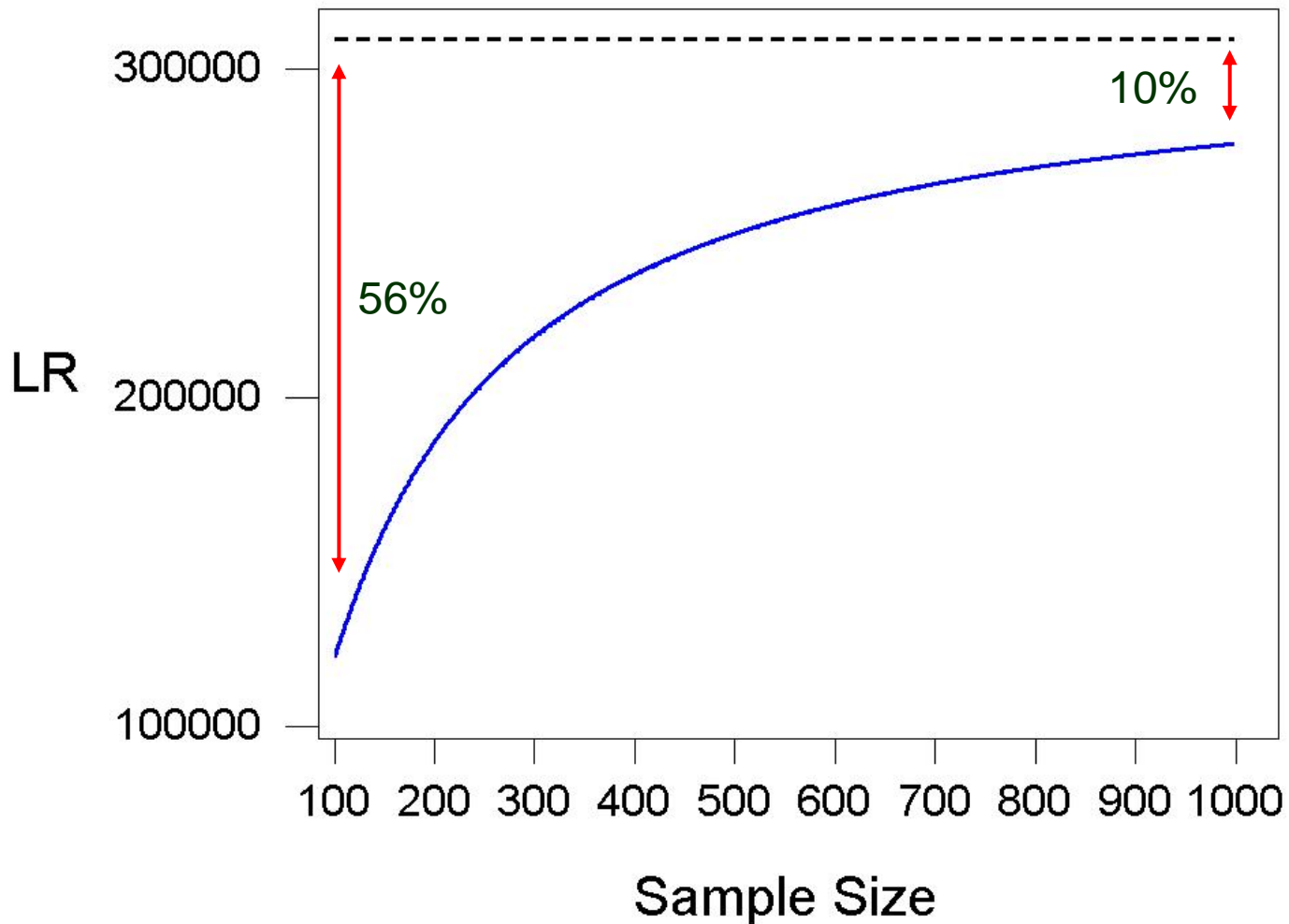
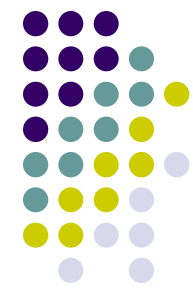
H_p : the victim and the suspect were contributors

H_d : the victim and one unknown were contributors

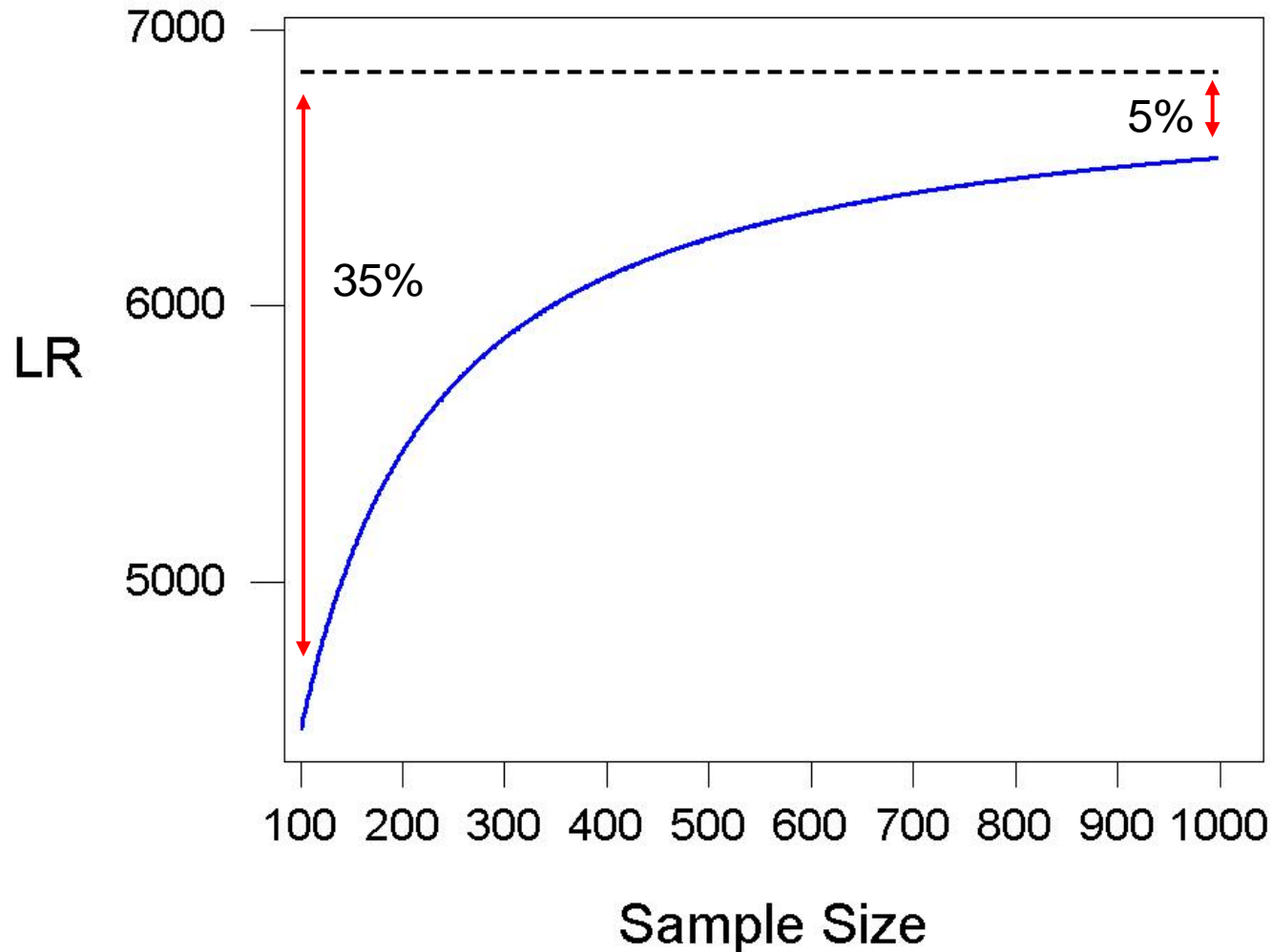
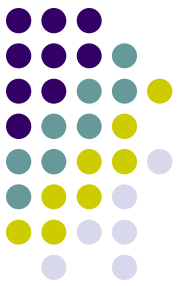
$H_{d'}$: the victim and a relative of the suspect were contributors

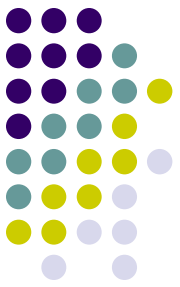
- Illustration: hypothetical database D with fixed allele frequencies and different sample sizes

LR of H_p : the victim and the suspect are contributors
vs H_d : the victim and one unknown are contributors



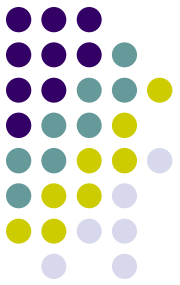
LR of H_p : the victim and the suspect are contributors
vs H_d : the victim and the cousin of the suspect are contributors





Summary

- Bayesian approach provides conservative evaluation of DNA mixtures
- Can be implemented efficiently by modifying existing plug-in formulae
- Incorporate subpopulation models to handle cases involving different ethnic groups
- Consideration of allele drop-out



Thank You!